

Towards a greenhouse treaty?

Last week's meeting of the Intergovernmental Panel on Climate Change at Stockholm seems to have done a useful job, the disappointments of some of its well-wishers notwithstanding.

MANY of the most interested observers, especially the environmentalist pressure groups, were disappointed by what they heard and saw at last week's meeting of the International Panel on Climate Change (IPCC) at Stockholm, but that was inevitable. Some had gone expecting a firm prescription of future emission limits for greenhouse gases, but found instead merely a revision of three documents in circulation since the early summer, some of which appeared to have been watered down. But it could not have been otherwise. IPCC is not an international authority empowered to issue global edicts, but an advisory committee. Its immediate function is to prepare a technical report for an intergovernmental conference arranged by the United Nations at Geneva in November. That conference will be asked to decide whether there should be an international convention on greenhouse gases. IPCC's report should be influential, but not decisively so; much more will hang on the inclinations of the participating governments. There is no other mechanism for making international law.

To the extent that some disappointment stems from the greater emphasis in the latest version of IPCC's report on the uncertainties in the greenhouse calculations, it is entirely misplaced. For there are uncertainties, not merely imprecisions of calculation but uncertainties about the physics of the problem. What is the fate of oceanic carbon and the effect of real clouds on ground-level insolation, for example? If IPCC were to make light of these conundrums, its report would be more open to just criticism and thus less persuasive of the doubters. There is even something in the view that governments fearful that the whole greenhouse rumpus is a device for meddling in their economic affairs will be more tempted to negotiate seriously if they are told there may be a problem — not that there is a problem.

That is why the best hope for November is that governments will agree to begin negotiating a convention on greenhouse gases. How should they do that? The first task is to produce a working list of topics to be covered by a convention. IPCC now includes a reference to the need for monitoring emissions, but not every participant in November will agree. And should there be penalties for proven transgressions? Should there be a procedure for amending the convention, and how will that be done? Who will decide how the damage done by different greenhouse gases should be balanced? Should the poor coun-

tries of the world be compensated for compliance, and if so how, and on what scale? What balance will there be between adaptation to climate change and its avoidance? And how (or by whom) will that be decided?

Even the core clause of a convention — an undertaking by the signatories that, from a certain date, they will hold emissions of greenhouse gases to prescribed amounts — leaves ample room for argument. Should limits be related to present emissions (thus penalizing developing countries) or to Gross Domestic Product (GDP, thus penalizing both poor countries and inefficient energy-users)? There would be merit in emission limits related to GDP per head of population, for then most poor countries would for some time be unconstrained while large but inefficient energy consumers would be tightly squeezed, but such a scheme would provoke cries that rapid population growth is a way of busting the convention.

The last difficulty shows that the greenhouse problem is not an isolated problem — one that can be 'solved' in isolation. Its links with the international monetary problem are even stronger. And while lists of questions yet to be answered may suggest that the negotiation of a treaty is an insuperable task, the truth is that they (and other questions) will have to be answered before a convention will stick. That is why there is a need to begin negotiations now, however woolly the draft treaty may be. IPCC's uncertainties will be mostly cleared up long before a convention is in place.

Completing them will be a long and difficult task unlikely to be finished this century. Many of IPCC's doubts should have been cleared up by then, even if others replace them. What the enthusiasts for a convention should grasp is that their goal will require of its signatories an unprecedented degree of self-denial. □

Remote research

New Zealand should battle imaginatively against the ill-luck of malevolent geography.

How should a small country make sensible policies for research and higher education? There are models of success, such as Switzerland and the Benelux countries, and of indifference, such as Ireland. New Zealand, by general consent one of the most delectable places on the



surface of the Earth with a population barely a tenth that of California, seems to be making heavy weather of the process. Will the well-intentioned plans now being hatched in Wellington yield what their authors intend, or the opposite?

The occasion for change has nothing to do with science. New Zealand, one of the richest countries in the world in the years immediately after the Second World War by virtue of its agricultural productivity, has since slumped from grace. The explanation is the rapid post-war growth of agricultural production in the United States and, more recently, Europe, followed by protection and increased competition from subsidized foodstuffs elsewhere in the world. So New Zealand is naturally an enthusiastic member of the Cairns group of countries battling in the current (Uruguay) round of negotiations in the General Agreement on Tariffs and Trade (GATT) for the liberalization of agricultural trade. But through much of this period, New Zealand has stuck fast to its own model of the welfare state; excellent (and free) health and education services and a general assumption that the government exists to protect people from ill-luck.

Imperatives

It is ironical that it is a Labour government that set about unscrambling this omelette, but Mr David Lange, elected prime minister in 1982, had no choice. New Zealand's overseas debt (per capita, greater than that of Brazil) was one imperative. The sense that New Zealand was stuck in some earlier decade (the 1950s, perhaps even the 1930s) was another. So, for eight years, there has been a determined effort to shrink the government's influence on affairs. Publicly owned services (airlines, broadcasting stations and telecommunications services) have been sold. Direct (income) taxes have been reduced in favour of a sales tax. The Labour government's intention is to concentrate on making policy, shedding as much public administration as possible to others — preferably private persons or corporations, otherwise (a second best) independent statutory corporations.

But this cannot wash for the infrastructure of the research enterprise, which in New Zealand (whence Rutherford sprang) is a model of over-achievement: success generally outstrips available resources and geographical isolation. A report last week (*Nature* **346**, 789; 1990) recorded how the government plans to make dependent research organizations compete for (or "contest") funds from a single pot of money while reserving for itself decisions about priorities. New Zealand's seven universities and its network of polytechnics have meanwhile been made directly dependent (with the abolition of the University Grants Committee last December) on the government, and are for the first time required to charge students tuition fees. The universities are now under pressure to contribute to the national research pool not only the NZ\$5 million for research administered by the old regime, but up to NZ\$20 million from their recurrent budgets for the doubtful privilege of

competing for a share of the larger pot within the framework of the government's priorities.

What, in these circumstances, should New Zealand be about? Research support for the major industries (meat, wool and forestry) is recognized as a priority (and the kiwi fruit, whose seeds now lodge between the teeth of the whole world's population, is one result). But it is not sufficiently appreciated that even the tradition of over-achievement will not, without extra funds, secure a cutting edge in fields such as artificial genetic manipulation. The big battalions elsewhere are already on the march. And, sadly, there is no sign that either the present government or the pollsters' bet as its successor at the general election on 26 October will see how urgent is the need for research training; skill is what will matter in the end, and graduate education is not the effete luxury it is too often considered in New Zealand.

Questions

There are other questions, most often unspoken. From most other places, New Zealand too often seems a remote island (but there are two of them). From there, it too often appears self-contained. But that view promises only an extrapolation of the present into the future, when young people will be driven away by the high cost of paying retirement pensions to the elderly. The real trouble is the twin handicap of geography and of a population (3.7 million) too small, in the modern world, to permit the efficient operation of Adam Smith's beneficent division of labour. One remedy is more immigration. The present government promises to push the rate of recruitment to 50,000 a year, but there have been recent years when even such an influx has been outweighed by departures. (Nobody seems bent on persuading Hong Kong Chinese to New Zealand between now and 1987.) Another is an Australian link, with a deepening of present economic arrangements in the short term, perhaps a confederation later on. But given the prevailing gradualism, neither remedy will suffice.

The real need, for New Zealand as a whole, is for a means of conquering geography. It is not sufficiently well-appreciated that the research and academic communities could pave the way for such a development. It is not merely that the global village has arrived, but that air fares are still a declining proportion of other costs and that the intellectual professions are among the most mobile of all. New Zealanders are keen that young people in research should visit overseas, but most arrangements for visiting overseas (invaluable though they are) are over-formal and rigid. Would it not make sense to devote some of the energy now being spent on the design of a grand new policy for national science on forging new research partnerships with people and laboratories elsewhere? And why should not universities in New Zealand set out to make joint appointments of able people with universities overseas? The people, after all, are as articulate as they come. It would be a shame if their future were defeated by mere geography. □

Assumptions of AIDS inquiry challenged

- Meaning of viral similarity questioned
- Contamination blunts effect of new results

London & Washington

AS the National Institutes of Health (NIH) inquiry into the part played by Robert Gallo's laboratory in the discovery of the AIDS virus enters its tenth month, astonishing new evidence has appeared in print that strikes at the central assumption behind allegations of misconduct.

Studies published in the August issue of the *Journal of Virology* imply that two groups of researchers have isolated almost identical strains of human immunodeficiency virus (HIV) from two totally unrelated patients — a finding that could undermine allegations that the close similarity between the virus isolated by Gallo's laboratory at the US National Cancer Institute and that discovered at the Pasteur Institute in Paris by Luc Montagnier must mean that Gallo's virus ultimately came from samples provided by Montagnier.

A race is now on to check the new results, with hectic efforts being made to trace the patients from whom the original samples were taken. Although a final verdict is unlikely before the end of the month, there are grounds for believing that the new results must at this stage be assessed agnostically. In particular, it is possible that both isolates are actually laboratory contaminants of the original French virus. If that proves to be the case, their similarity would be further evidence of the ease with which HIV can spread accidentally in the laboratory.

The viruses were characterized by David Volsky's group at Columbia University in New York and Mario Stevenson's laboratory at Nebraska University. As reported, they come from two independent isolates, one obtained in 1984 from an AIDS patient in New York.

ASTRONOMY

Mt Graham go ahead

Washington

AFTER two years of court battles, protests and federal investigations, Arizona's Mount Graham may finally have a telescope. The US Forest Service announced at the end of last month that US conservation law does not require further studies of the project and issued a building permit on 24 August. Construction of the first of three telescopes on the 10,700-foot peak, which is the only home of the endangered Mount Graham red squirrel, will begin within weeks, according to a spokesman at the University of Arizona.

Christopher Anderson

the other from a seropositive haemophilia patient in Nebraska in 1989. An analysis by Gerald Meyers of the HIV sequence database at Los Alamos (see News and Views page 18), shows that they differ by less than 2 per cent. This finding jars with the conventional view that viral isolates from different individuals, or even the same individual at different times, are bound to display marked genetic differences.

Significantly, the sequences are not only conspicuously similar to each other, but also to those of two other isolates, LAV1 and HTLV-IIIB — the names given to the original viruses discovered by Montagnier and Gallo in 1983 and 1984 respectively.

The panel conducting the inquiry at NIH's Office of Scientific Integrity (OSI) has spent much of the summer quizzing AIDS researchers and poring over notebooks from Gallo's laboratory. High on their agenda has been the issue of whether or not viruses as similar as HTLV-IIIB and LAV1 can emerge from two unrelated patients. On the face of it, the genetic sequences reported in the *Journal of Virology* articles say that they can.

Deputy OSI director Suzanne Hadley says that the inquiry panel is well aware of the implications of the sequences, but added that, although the genetic similarity of different HIV isolates "is an important issue, it's not the only issue". When asked if the panel would be commissioning an independent study to verify the sequences, she said, "That kind of analysis may be an interesting thing to do. There's no area closed."

But there are signs that the new sequences may not be quite what they seem. In particular, a process of contamination cannot be ruled out, especially because Stevenson worked as a postdoctoral researcher in Volsky's laboratory at the University of Nebraska. Before Volsky moved to New York in 1987, he helped to establish the cell line used to grow the viral isolate from the New York patient, termed N1. It was from this cell line that, in December 1984, Volsky derived the cloned virus N1T, which he used in later biological studies and finally sequenced.

The virus sequenced by Stevenson's group in 1989 is known as MFA. According to their *Journal of Virology* paper, it was derived from an isolate called MF, which was originally obtained from a clinic in Nebraska. Nonetheless, when asked about the possibility of contamination, Stevenson made this statement: "We

maintain the source of the original N1T isolate, N1-infected CEM cells, in our laboratory, and have cited that cell line whenever we've used it. Given the similarity of the MFA and N1T sequences to HXB2 (HTLV-IIIB) and BRU (LAV), I must accept the possibility that the MF isolate was in fact derived from cultures contaminated with N1-CEM cells."

But even if Stevenson's MF virus does come from a contaminant, the remarkable similarity between Volsky's N1 isolate and LAV1 remains. Could the cells used by Volsky's laboratory to culture the original N1 isolate somehow have become contaminated with LAV1?

Volsky says that between 1984 and 1985 his laboratory prepared about 1,000 different viral cultures using only HIV-infected serum from a clinic in New York. During the autumn of 1984, however, he says that the original cells containing the N1 isolate were co-cultivated with CEM cells donated by Montagnier's laboratory.

Volsky has maintained that, to the best of his knowledge, these were virus-free and would have been routinely checked for signs of virus before being added to N1 cultures. But Montagnier's laboratory records suggest that, in 1984, he sent Volsky LAV1-infected CEM cells, in addition to any virus-free cells of which there is no specific mention in the records. "He asked for LAV-infected cells, and we gave him those", says Montagnier.

Although Volsky cannot recall receiving LAV1-infected cells, he admits that in the light of Montagnier's record they were probably sent. "If we received contaminated CEM cells harbouring latent virus that became reactivated during co-culture, there is nothing one can do", says Volsky.

Regardless of whether infected cells were actually received, Volsky stresses that an active culture producing LAV1 was never established in his laboratory.

William Blattner, an epidemiologist at the National Cancer Institute, has been searching for Stevenson's patient so as to resequence the virus. Although he has no name for the patient — nor does he know if he or she is still alive — the number of possible candidates has been narrowed to 20 or 30, he says.

By contrast, Volsky's patient has been traced and is still alive. Since virus was first isolated from him in 1984, he has become seronegative by the standard antibody test. Nevertheless, Volsky's group has managed to detect HIV genetic material in his blood and now plans to sequence it. They also intend to sequence DNA from frozen samples of the N1 isolate obtained in 1984. The sequencing should be completed by December, says Volsky, who believes that final judgment on the origin of the similarity of N1T and LAV1 should be reserved until then.

David Concar & Christopher Anderson

Will ban provoke challenge?

Washington

THE board of directors of the Parkinson's Disease Foundation (PDF) will decide on 11 September whether or not to sue the US administration over its ban on the use of federal funds to support fetal tissue transplantation research.

But although patients with Parkinson's disease are one of the main groups that could potentially benefit from a lifting of the moratorium, Dinah Orr, PDF executive director, says, "we will not want to take the whole thing on our own shoulders". So far, the United Parkinson Foundation and the Association of American Medical Colleges (AAMC), which represents medical schools, teaching hospitals and academic societies, are also considering whether to support the legal challenge.

A temporary ban on research involving the transplantation of human fetal tissue derived from induced abortions into human recipients was introduced in March 1988 by the then Assistant Secretary of Health Robert Windom. It was precipitated by a request by scientists at the National Institutes of Health (NIH) to carry out an experiment involving the transplantation of human fetal cells into the brain of a patient with Parkinson's disease. Windom withheld approval, but convened an NIH panel of ethical, legal

BRAZILIAN SPACE PROGRAMME

Iraqi link threatens progress

São Paulo

BRAZIL'S space programme is one of the victims of the crisis in the Middle East triggered by Iraq's invasion of Kuwait — or so at least the Brazilian Air Force believes. The air force has responsibility for the development of Brazil's Satellite Launcher Vehicle and claims that the United States will not return parts of the rocket sent for thermal treatment because of Brazil's military links with Iraq.

In the past, Brazil has sold much military equipment to Iraq, including rockets developed from scientific sounding rockets. It also has experts in Iraq helping with a programme to develop an air-to-air missile, the Piranha, which is similar to the US Sidewinder missile.

The United States has long been suspicious that the satellite launch vehicle might be converted into a long-range missile, and is concerned that the project is managed by the Brazilian Air Force rather than a civilian agency. But no immediate military application is possible as the rocket is not expected to be ready until 1994 — five years behind the original schedule.

Ricardo Bonalume

and scientific experts to examine the issues relating to human fetal tissue transplantation research.

Although the panel concluded that such research should be permitted and the ban on federal funding lifted, Health and Human Services (HHS) Secretary Louis Sullivan rejected the advice and extended the ban indefinitely in November 1989 (see *Nature* 342, 105; 1989).

An attempt by Representative Henry Waxman (Democrat, California) to overturn the ban by legislative means will have an uphill struggle when his bill is considered by the health and environment subcommittee in early September (see *Nature* 346, 598; 16 August 1990).

David Moore, assistant director in the Office of Government Relations at the AAMC, says his organization supports Waxman's initiative, but believes that the chances of the legislation passing this term "are pretty remote", which is why AAMC is keeping all its "options open". The AAMC's executive council meets at the end of September to consider whether or not to support action to challenge the ban in the federal court, by which time, Moore says, they should have a better idea "whether or not the legislation is going to go forward".

A legal study by the law firm of Fulbright and Jaworski, commissioned by AAMC, states that "the merits of a lawsuit challenging the moratorium appear strong" and could proceed on two fronts. First, the HHS failed to comply with the Administrative Procedure Act, which requires both public notice and an opportunity for public comment before a rule is established. Second, the study says that the Secretary's action appears to be "arbitrary and capricious", in that the only rationale given for extending the ban indefinitely was that it might cause an increase in the number of abortions.

"Instead of making it a factor to consider in weighing whether it was appropriate to fund the research, he [the Secretary] made it the determining factor — the only factor", says attorney Stephen McNabb who helped prepare the study.

Michael Astrue, General Counsel for the Public Health Service (PHS), is unconcerned by the prospect of a lawsuit and argues that the PHS statute "clearly commits discretion to the Secretary to either engage in this research or not". By making the ban indefinite and not permanent, Astrue says "the Secretary wanted to indicate that he would continue to have an open mind on the issue", particularly as the abortion law is in a state of flux and because technology in this area changes very rapidly.

Even if the Parkinson's disease organizations and AAMC decide to go ahead,

More (students) means more (pain)

London

BRITISH higher education will become "a demoralized, poor quality system" unless there are increased funds for teaching to support the government's planned expansion of student numbers, warn university vice-chancellors and polytechnic directors. But a large increase from public funds is unlikely, whatever the political colour of the next government, so increased costs may have to be borne by students.

A report from a working group set up by the Committee of Vice-Chancellors and Principals and the Committee of Directors of Polytechnics (the first occasion universities and polytechnics have acted together in this way) suggests a number of financial mechanisms, all involving increased payments by students. One possibility is for students to borrow money to meet some of their teaching costs, similar to the controversial 'top-up' loans scheme under which, starting this academic year, students will over the next few years borrow to cover an increasing proportion of their living expenses. Other suggestions include the introduction of a 'graduate tax' and charging teaching costs in full directly to students, with government support provided by a system of scholarships. Working group chairman Professor John Ashworth, from the University of Salford, says that the report's primary importance is not to produce a detailed funding mechanism, but rather to combat the government's current 'wait and see' policy on the finances of higher education. "A decision will have to be made this year", he says.

In the run-up to a general election, further upheaval of higher education may not be a government priority. But the costs of teaching extra students are expected to produce a crisis in 1991-92. Under the new competitive bidding system for student places, universities have asked for a 19 per cent increase in students over the four years to 1994-95. But most universities have bid for places at, or only just below, the Universities Funding Council's suggested maximum guide prices (see *Nature*, 345, 468; 7 June 1990), subverting the government's plan to realize 'economies of scale', increasing student numbers at a lower average cost.

Peter Aldhous

important issues remain. They will need to muster support from other disease groups to meet the legal costs, and will have to decide and which of them, from a tactical point of view, should take the lead. As Lawrence Hoffheimer, attorney with Noto, Oswald, Hoffheimer, Eisman and Miller, puts it, "who has standing [the legal term for the right of a plaintiff to raise the claim] and who pays for the suit are two different matters".

Diane Gershon

TROPICAL ECOLOGY

The view from the top

Washington

JUDGING from their first few weeks of work, researchers at the Smithsonian Tropical Research Institute are confident that the tower crane they have set up in Panama can conquer tropical ecology's last frontier—the tree canopy where most of the photosynthesis takes place and a diverse community of largely unstudied plants and animals lives.

The experimental crane is not actually in the midst of the jungle, as might appear from the picture here, but in a small (200 hectare) medium-age forest in Panama City's Metropolitan Nature Park. Nor does the crane belong to the Smithsonian; it has been rented until next April from a Panama City contractor. But once researchers are satisfied that the crane can

AIDS

Commission grows angrier

Washington

OVER-rigid ground rules for participation in clinical trials of AIDS patients are excluding large segments of the infected community, including blacks, women, children and drug users, a US congressional advisory panel concluded last week. The National Commission on AIDS, in its third and hardest-hitting report to the president, said that the 12,000 people participating in AIDS trials reflect a "pitifully small" fraction of the those eligible. Blacks and Hispanics account for about 43 per cent of all AIDS cases, but only 23 per cent of the clinical trial populations, the panel noted.

The panel also criticized the National Institutes of Health (NIH) for delay in expanding their research to include opportunistic infections, the leading cause of death for people with AIDS. It noted that although \$428 million has been spent by the US AIDS Clinical Trial Programme, most approved drugs for AIDS and AIDS-related infections have been developed outside the programme. NIH officials say efforts are being made to increase minority participation in clinical trials and the research emphasis on opportunistic infections.

Other concerns included the increasing incidence of AIDS in rural communities. In one year, there has been a 37 per cent increase in reported AIDS cases in rural areas compared with a 5 per cent rise in metropolitan areas. The panel attributes much of the growth in rural cases, especially among heterosexual women, to "the combination of crack cocaine, trading sex with multiple partners for drugs or money, and the rising rate of syphilis infections". Christopher Anderson

provide secure and repeated access to the forest canopy anywhere within reach of its 36 metres span, they hope that they will be able to persuade the US Congress to provide funds to buy and set up a larger crane at their main Panamanian research site on Barro Colorado Island.



Treetop laboratory — experimental crane for gathering data in tropical forests.

Alan Smith, assistant director for terrestrial Research at the Smithsonian Tropical Research Institute, who heads the crane project, says his team is "delighted at the rate at which we are getting data, and the diversity of the data". Researchers ride in a special gondola which delivers them to any designated point in the canopy. Once there, they hook onto a branch to stop the gondola swaying too much, and get down to observations. Studies of photosynthesis, transpiration and

insect diversity are under way. A key part of future projects will be to see how rates of photosynthesis are affected by the changes that might occur as greenhouse gases build up and global warming occurs; without this information it will be very hard to predict how climate change will affect tropical forests.

Smith believes the crane will provide more convenient access to the canopy

than the time-consuming and dangerous systems of towers and aerial ropeways now in use. One earlier concern seems to have been unfounded: operation of the crane does not appear to disturb animals in the region. If the Smithsonian gets its own crane next year, it will be moved from one location to another by helicopter so that no damage is done on the ground. A crane with an 80-metre arm is favoured that will give access to five hectares of forest canopy.

Alun Anderson

ACADEMIC INSTITUTIONS

Thanks, but no thanks for impact theories

Washington

ALLAN Kelly, an 89-year-old California cattle rancher who believes that asteroid impacts are responsible for the Earth's tilt, the end of the last Ice Age and the biblical flood, has had some trouble in persuading the scientific community to see things his way. So earlier this year he offered California State University at San Marcos (CSU) \$250,000 to set up an endowed chair for a geologist to research his theories for him. According to the proposed agreement, the scientist who holds the position will "study . . . the truth and falsity of Mr Kelly's theories", reporting back to Kelly annually on the progress of the research.

The offer stipulates that the chair holder be named the "premier professor" of arts and sciences at the university.

On 28 August, CSU accepted the offer. University president William Stacy told the *Los Angeles Times* that the agreement

protected the academic freedom of the chair-holder, while also giving Kelly "some reasonable assurance that the matters of special concern will receive scientific consideration". But the next day, after the *Los Angeles Times* published an article predicting ridicule and quoting geologists who said that no serious scientist would take the chair, the university quickly reversed its decision and returned the money to Kelly.

Although university geologists were not opposed to an endowed chair in the general area of impact theory, they were appalled that CSU would allow a donor to dictate the specific details of the research. After seeing the *Times* article last week, Stacy decided that other faculty members would probably feel the same way. "It's crucial to avoid any further embarrassment to the university", he said in announcing the hasty change in plans.

Christopher Anderson

Sex, racism and videotape

Washington

A Canadian psychologist whose theories on the links between human intelligence, race and sexual appetite have drawn wide protest will have to teach most of his classes by videotape next term to avoid violent clashes with demonstrators.

The University of Western Ontario officials announced last month that Philippe Rushton, a tenured professor in the psychology department, will teach all his undergraduate courses "using a personalized system of instruction (PSI) format". What that means, the university said, is that Rushton will videotape a 90-minute weekly lecture, which will then be available for viewing throughout the week at the convenience of his students.

Protesters, of whom the university expects hundreds, will find no class to disrupt or demonstrate outside.

Rushton has protested at the decision as an abridgement of his academic freedom, and has formally appealed against it. But university officials say that their decision was actually the only way to preserve that freedom. "We are expecting a concerted, highly organized protest on the first day of class [10 September]. We decided that we must protect [Rushton's] freedom to teach, and this was the way to do it", says university provost Thomas Collins.

Since Rushton's theories were first publicized in the press in 1989, he has been the focus of a storm of controversy. The premier of Ontario has publicly stated that Rushton's views are "morally offensive to the way Ontario thinks", and Rushton has been investigated by a special police force concerned with pornography and "hate literature" (its findings: "loony, but not criminal").

His most controversial theories revolve around his attempts to measure behavioural and physiological differences between races and to relate them to past evolutionary history. Rushton has compared more than 60 different variables — including such factors as cranial capacity and genital size — and concluded that a continuum can be found, with Mongoloids at one end, Negroids at the other, and Caucasians somewhere in between. His work is partly supported by the Pioneer Fund, a New York-based private organization that has been widely attacked for funding research related to eugenics. Rushton's appearance on "Geraldo", a US television talk show, and massive press coverage in Canada have helped to stimulate calls for his dismissal and demonstrations outside his classrooms.

Last year, Rushton circulated a letter appealing for support from other researchers, which brought in over 40 letters, some of them from well-respected scientists. Although many disagreed with

Rushton's views, there was considerable support for his right to pursue his own ideas and protest over the "unsatisfactory" rating the university had given Rushton for his recent academic performance. (Three "unsatisfactory" ratings can lead to the dismissal of even a tenured professor.)

Until last month, the university and Rushton were negotiating a strategy for the new term to avoid more outbreaks of violence. Although tentative agreement had been reached on a plan to allow Rushton to teach in a portable classroom (which could be ringed by police and isolated from other classes) near the perimeter of the campus, negotiations broke down over Rushton's demands for security. "He said that if we had 300 demonstrators, we should provide 300 security guards. Obviously, we can't do that," says Collins. The university solution — videotape — has been received no better by Rushton, but outside support is proving difficult to muster.

The Canadian Association of University Teachers has asked a panel to address the issue, but executive director Donald Savage says that the requirement that professors videotape lessons is neither unprecedented nor necessarily a violation of academic freedom. In large introductory classes where classroom space is a concern, the practice is quite

ARCTIC RESEARCH

The cold war ends at Resolute Bay

Washington

REPRESENTATIVES of all eight countries with territory in the Arctic agreed last week to establish a new nongovernmental International Arctic Science Committee (IASC) to coordinate research in the



region. The agreement, signed at the remote scientific outpost at Resolute Bay, has been several years in the making and only became possible with the coming of *perestroika*.

The region is still ringed with radar early-warning stations, air bases (commanded on the western side from Thule

Longest running reactor retires

London

THE world's longest continuously operating nuclear reactor was shut down on 3 September. The Graphite Low Energy Experimental Pile (GLEEP) has been in service at the UK Atomic Energy Authority's Harwell laboratory for the past 43 years. Designed for the study of reactor physics, it demonstrated the stability of graphite-moderated gas-cooled reactors, which went on to form the backbone of Britain's nuclear electricity generating industry.

For most of its life, GLEEP ran at a fraction of its potential power, as a versatile research reactor, used for testing reactor materials and calibrating radiation monitors. Peter Aldhous

common at some universities, he notes.

The university faculty association has not taken a formal stand on the issue, preferring to wait for the conclusion of the university grievance process. Some other faculty members privately say that they are troubled by the university decision, but few have rallied to oppose it. "As unpalatable as I find this solution, it avoids the possibility of flashpoints of violence, and it ensures that we don't have to stop [Rushton's] classes," says psychology chairman Greg Moran.

Christopher Anderson

in Greenland) and stations for submarines which operate under the ice pack.

Canada, Denmark, Finland, Iceland, Norway, Sweden, the United States and the Soviet Union now invite other countries with Arctic research programmes but without territory in the region to seek membership in IASC. The United Kingdom, France, Germany and Japan are thought likely to be the first to join.

The greatest advantage conferred by membership will be the opportunity to coordinate research and share resources in a region where logistical problems are often overwhelming. The Soviet Union has a large fleet of powerful ice breakers, Canada has several research ships and the United States has good satellite coverage of the area — put together, the whole should permit research programmes that are not now possible.

Research in the Arctic will become increasingly important as the first signs of global change appear. Greenhouse warming is expected to be greatest at the poles and to lead to changes in the sea ice cover and the melting of glaciers and ice sheets.

Alun Anderson

AIRPORT CONSTRUCTION

Ocean mud on Kansai's face

Osaka

JAPAN'S government as well as its civil engineering contractors are acutely embarrassed that the artificial island meant as the site for the new Kansai International Airport (KIA) here is sinking into the sea at a rate of 5 cm a month.

Foreign construction companies were turned away from the huge construction project on the grounds (among others) that they lack experience of building on the soft marine sediments around Japan. The same explanation was offered in 1987

when access to the project be-

metres with pillars of sand before the reclamation operation began early last year.

Three mountains are being excavated on the adjoining mainland and on nearby Awaji Island to fill the site. Every day, 80 huge barges each dump 3,000 cubic metres of rock and rubble at the site in a computer-controlled operation that runs from 4 a.m. in the morning to 11 p.m. at night.

By March this year, it was clear that the emerging island is sinking much faster than the KIA company had predicted when making budget estimates for the project. At present the island is sinking

about 5 cm a month and the latest estimates suggest that the total subsidence many years after completion will be about 9–11 metres compared with the 6–8 metres originally estimated.

The problem is not with the soft surface muds of which Japanese civil engineers claim exclusive knowledge — subsidence of the 20-metre-thick surface layer of clay is likely to be close to the 5 metres or so predicted by the KIA company. But the company's estimate of subsidence of the underlying Pleistocene clay and sand layers seems way out. The company predicted subsidence of only 1–2 metres, but actual subsidence is likely to be 5–6 metres according to

the latest measurements.

Ironically, Akio Nakase, emeritus professor of Tokyo Institute of Technology and one of the key advisers for the project, came close to predicting the subsidence correctly in 1987 (see figure), but KIA officials decided to use smaller estimates based on a simpler model because the projected costs of reclamation would be less.

There are two reasons for the discrepancy between the company's estimate and Nakase's 1987 estimate. The Pleistocene layers contain many thin lenses of sand which act as drainage channels for ground water squeezed laterally out of the sediment as the rapidly growing island presses down with a force of about 40 tonnes per square metre. The greater the number of layers, the greater and more rapid the subsidence. The company's estimate assumed only two drainage layers thicker than 3 metres and ignored all thinner layers, while Nakase's calculations included seven drainage layers as little as 1 metre thick.

A second problem is that of estimating the progress of compression of the consolidated Pleistocene sediments. At a certain critical pressure, the bonds be-

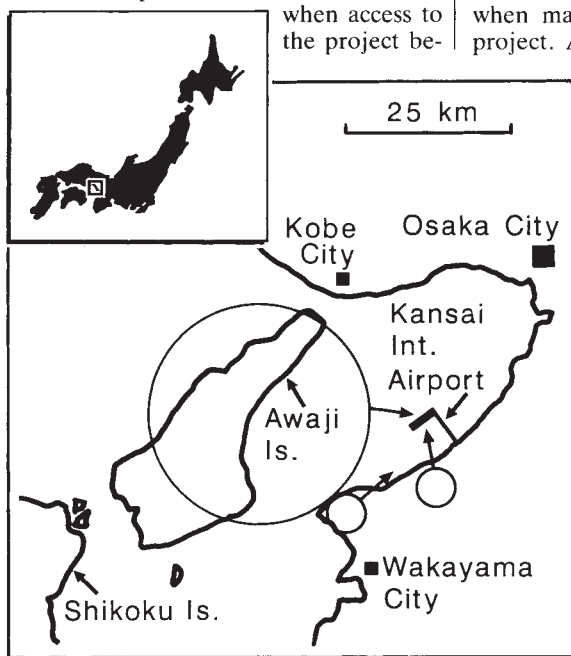
tween sedimentary particles break and compression becomes much more rapid. The classical model of compression used by the company has proved inadequate, but one developed by KIA's group of geotechnical advisers seems to be more accurate.

Despite the subsidence, KIA officials are determined to open the airport on schedule on 31 March 1993; the company has installed a clock at the site showing the number of days to completion. But because the subsidence will then still be progressing at a fast and uneven rate, the Italian-designed airport terminal building will have to be equipped with compensating hydraulic jacks in its foundations.

Nakase's personal opinion is that the opening of one of the two wings of the terminal (the south wing) should be delayed for two or three years, when the number of hydraulic jacks could be greatly reduced. But KIA officials insist that the airport must be complete in 1993 because the present Osaka international airport cannot cope with the rapid growth in passengers and air cargo and because the company will then have a huge debt of about ¥700,000 million, and the interest payments to go with it.

More generally, Japan is now being criticized for the inadequacy of its international airport facilities, which are effectively shutting foreign airlines out of Japan's lucrative air market.

David Swinbanks



came a major bone of contention in a trade dispute with the United States, Europe and South Korea. The government will now be embarrassed that the island under construction is sinking into soft underlying sediments much faster and further than predicted by the Japanese contractor.

Geotechnical experts and officials of KIA Company Ltd met on 1 September to make more accurate predictions of the subsidence. Details of the meeting have not been made public, but it is clear that several extra metres of fill will have to be added to the emerging island, at a cost of more than \$100 million, if the airport site is to be sufficiently above sea level.

The reclamation project is of unprecedented scale and involves building an island more than 4 km long and 1.25 km wide in 18–20 metres of water 5 km off the coast. The aim of building at sea was to avoid the problem of aircraft noise in residential areas, so that the airport can operate 24 hours a day.

Work began in 1987 with the construction of a sea wall more than 11 km long around the site. The soft seabed inside the wall was reinforced to a depth of several

First mud and then mortars

Osaka

Subsidence is not the only problem besetting Kansai International Airport. Last Friday, unprovoked by this correspondent's invitation to the site, radicals opposed to the airport launched a mortar attack on one of the excavation sites for fill. At 3 p.m., nine hours after the mortar attack in the early hours of the morning, a fire was still raging in a nearby forest and a pall of smoke hung over the airport construction site. Eighteen fire engines were needed to put out the blaze.

A man claiming to belong to the radical ultra-leftist organization Chukakuha (middle core faction) claimed responsibility for the attack. Chukakuha has violently opposed the Tokyo International Airport at Narita, making it necessary for the airport to be surrounded by barbed wire and armed guards.

Chukakuha claims to support farmers who were opposed to giving up their land for the airport at Narita. But it is not clear why it opposes the new Kansai airport except that it is calling it the "Narita of the west".

David Swinbanks

Hankering after past glory

Berlin

WITH the approaching reunification of Germany, Berlin is hoping to assert itself again as one of world's great centres of academic excellence, on a par with Boston or Berkeley. But there are many obstacles in the way and it is by no means clear how and when they will be overcome.

Until the Nazi takeover in 1933, Berlin could claim to be the preeminent scientific city in Europe, thanks to the benevolent policies of Kaiser Wilhelm II and the vigorous efforts of researchers such as Max Planck. Now the challenge is to recreate that strength by integrating the institutes created after the Second World War on the two sides of the Wall.

The reunified Berlin will contain three large universities, more than 20 West German research centres and approximately half of the nearly one hundred scientific institutions in East Germany. "At least quantitatively, we are on a par with the top cities of science in Europe and even in the United States", says Werner Vöth, vice-president of the Free University of Berlin.

For very different reasons, science in both West and East Berlin profited from the 40-year division resulting from the Cold War. West Germany jammed the city with institutions of science and culture to keep it vital and attractive despite its location in the middle of East Germany and the departure of most of its heavy industry in the 1950s and 1960s. And East Germany made East Berlin a showplace of scientific and technical prowess — to the detriment of the hinterlands.

The city has other advantages, principally its location. Once perched on the periphery of a divided Europe, Berlin now stands at the threshold of a revitalized Central and Eastern Europe. As in its glory days from 1880 to 1930, Berlin could once again become a focus of East–West exchange, especially if it maintains the contacts with Moscow established by East Germany.

But renewal will be arduous. Supporters of Berlin's renewal, led by West Berlin Science Senator Barbara Riedmüller-Seel, must negotiate a minefield before they reach their immediate goal of a plan for the integration and unification of science in the city.

The first task is to find out exactly what East German science has to offer. Despite the disappearance of the Berlin Wall, research in East Berlin is still an unknown

quantity. Over the objections of those who first wanted to cut off support to East German researchers after reunification (on the grounds that many researchers were compromised by being members of the Communist Party), the West German science advisory council Wissenschaftsrat will carry out a project-by-project evaluation of the grossly overstaffed East German Academy of Sciences, where the Communist regime concentrated its best researchers and equipment (see *Nature* 346, 209; 1990). Wissenschaftsrat will say whether institutes should be absorbed by universities or industry or be shut down. Politicians have promised to abide by the



Work is progressing on the reconstruction of the Reichstag building, gutted by fire in 1933 at the time the Nazis rose to power and wiped out parliamentary democracy in Germany.

recommendations.

Some 45 per cent of the 26,000 academy employees work in Berlin. It would not be unrealistic to estimate that half will remain after the Wissenschaftsrat recommendations are carried out, says Riedmüller-Seel. But only after the recommendations begin to emerge later this year can she incorporate in her planning the 6,000 people who are expected to remain.

After the future of the academy is settled, the next goal will be to raise the scientific quality and increase the number of students at the Humboldt University in East Berlin, which as recently as 1988 taught only 12,000 students (now 16,000) with 5,000 faculty and staff, a ratio of which West Berliners could only dream. Altogether, the universities and technical colleges of Berlin will have over 120,000 students, more than in any other city in Germany.

Although some reductions in the universities are planned due to redundancy (especially in smaller faculties such as Chinese or Japanese studies), the Free University and Humboldt University will both continue to offer courses in popular

fields in the humanities and social sciences, says Riedmüller-Seel, a policy necessary to meet the influx of East German students. Western students are not expected to enter the Humboldt University for several years, after the quality and image of the faculty (and its members' pay, now one-third of that in the West) has improved.

Perhaps the biggest problem that Riedmüller-Seel will have to overcome is the continuation of regional competition. As both East and West Berlin were heavily subsidized by their respective governments at the expense of other regions, reunification is bringing resentment of Berlin out into the open, especially in the south and west of West Germany. This is particularly dangerous because Berlin will

be a separate *Land* (state), one that is top-heavy with science. Riedmüller-Seel says she "expects a big fight" with the other *Länder* over aid to Berlin. The Bonn Research Ministry appears sensitive to Berlin's plight, claiming that support for Berlin is increasing.

Riedmüller-Seel realizes that it will require a clever and tenacious policy of faculty recruitment — a process that could take decades — before Berlin can raise its scientific standards. Though Vöth of the Free University asserts that no professor has turned down a job offer there since the opening of the Wall, he admits that there is no extra money in his budget to pursue the best researchers, especially in the natural sciences.

This is where the Max Planck Gesellschaft (MPG), West Germany's elite research organization, could help, says Riedmüller-Seel. MPG, which is based in Bavaria, is "fixated on southern Germany", she says, and that she will try to persuade MPG to consider placing more groups, if not entire institutes, in Berlin.

If there is to be a renaissance in Berlin, much will depend on how soon it is named the capital of Germany. Since the opening of the borders, many residents claim that the quality of life has deteriorated. Higher prices, the hordes of East German shoppers and the machinations of property agents are all causing complaint.

But once Berlin becomes the capital, there will be an influx of government agencies, foundations and even industrial lobbies that will help to invigorate the research scene, says Dietrich Mahlo, a member of the Bundestag (parliament) from West Berlin. Vöth also points out that what used to be East German villages beyond the wall are now low-rent suburbs easily accessible to the Free University.

Steven Dickman

IPCC NEGOTIATIONS

Trench warfare at Sundsvall

London

THE first step on the long road to an international agreement to combat global warming was completed last week in Sundsvall, Sweden, when delegates from 74 nations reached agreement on the final text of the report from the Intergovernmental Panel on Climate Change (IPCC). The report will be used to begin negotiations on an international climate change convention, similar to the Vienna Convention which provided the impetus for efforts to protect the ozone layer. But negotiations will be difficult, with a group of nations led by the United States opposing urgent action to limit greenhouse gas emissions.

Earlier in the meeting, Peter Timeon, foreign secretary of the tiny South Pacific nation of Kiribati, made an impassioned plea for stronger action from the industrialized nations. His country is one of those most at risk from the sea-level rise predicted if global warming continues unchecked. But his words apparently went unheeded by US, Soviet and Saudi Arabian officials, who resumed their painstaking battle over the report's precise wording with the European representatives who were supporting early action.

The report's important role in framing a climate change convention explains the almost absurdly slow debate over its contents, likened to "trench warfare" by one observer in Sundsvall. US delegates, concerned about the impact of a rigorous anti-global warming strategy on the US economy, fought strongly to introduce references to the scientific uncertainties

surrounding climate predictions and the fact that the economic costs of measures to combat the greenhouse effect have not been adequately calculated.

Environmentalists were angered by what they saw as US attempts to undermine the report. John Houghton, from the UK Meteorological Office, chairman of IPCC's working group I, concedes that people looking for a strong line from IPCC may be disappointed. But IPCC was set up to define the options in combating and adapting to climate change, not to produce a definitive political solution, Houghton says. He adds that the US delegation could have blocked the adoption of the report in Sundsvall, but did not do so.

The Sundsvall meeting produced a concluding overview to link the reports from IPCC's three working groups, which have looked at the science and impacts of climate change and responses to the problem. The individual working groups' reports (on which the overview is based) were finalized earlier this summer, but the official release of the impacts and responses reports was delayed until the Sundsvall meeting.

A leaked early draft of working group III's report, which examines responses to global warming, was attacked by environmentalist groups for the modest targets it suggested for future carbon dioxide emissions: the proposals for each region broadly paralleled the existing positions of governments (see *Nature* 345, 373; 31 May 1990). The final version has dropped the suggested emission targets, instead discussing in more general terms the

various strategies governments may adopt, from adapting to climate change through to energy efficiency measures and cuts in greenhouse gas emissions.

Fiona Wier, Friends of the Earth's delegate in Sundsvall, says the final report does little to quantify the reduction in carbon dioxide emissions that could be achieved very quickly through increased energy efficiency, arguing that its proposals "could have been written on the back of an envelope two years ago". But she is pleased that the specific, but modest, suggested emission targets have now been left out. During the lengthy negotiations towards a climate convention, individual countries may take unilateral action to cut their emissions. Wier believes these moves would have been restricted if group III's original draft had been incorporated into the final IPCC report.

Although IPCC's primary function was to produce the now-completed report, Houghton says that it will have an important role to play in the continuing climate convention negotiations, providing regular scientific and technical updates for the negotiating teams. **Peter Aldhous**

HUMAN GENOME PROJECT

Database goes on-line

London

GENETICISTS gather in Oxford this week to test a new computer database for the Human Genome Project. The database, developed by a team led by Richard Lucier from Johns Hopkins University, Baltimore, Maryland, should become available to genome mappers later this month.

The new database will contain basic gene-linkage data, information on polymorphism, and practical details on genetic probes and the location of laboratory material. The genetic sequences themselves are deposited in the Genbank database at the Los Alamos laboratory, but Lucier says that discussions to allow the exchange of data with Los Alamos are under way.

Similarly, the database will be linked up to the existing On-line Mendelian Inheritance in Man database, the leading repository for human genetic disease data (also at Johns Hopkins) and, it is hoped, the mouse genome mapping database at the Jackson Laboratory in Florida, Lucier says.

Initially, copies will be held on mainframe computers at Johns Hopkins, and at a site in the United Kingdom — probably the Medical Research Council's Clinical Research Centre in North London. But there are plans to extend the service to the West German Cancer Centre in Heidelberg and later to Japan and Sweden, so that geneticists in those countries will not have to depend on sometimes unreliable international electronic communications to access the genome data. **Peter Aldhous**

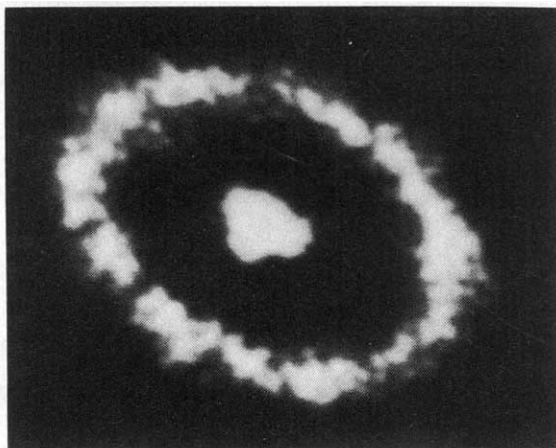
No blurry smudge, this Hubble photograph

NASA, the National Aeronautics and Space Administration, last week released its second batch of impressive stellar photographs from the Hubble Space Telescope to prove that the orbiting observatory is far from blind.

One image (right), taken with the European Space Agency's Faint Object Camera, shows a mysterious elliptical ring of glowing material around the remnants of Supernova 1987A.

NASA astronomers said that the ring is destined to be a relatively short-lived structure. Expanding debris from the explosion (now delineated by the blob at the centre of the ring) will overtake and disintegrate the ring within a century.

Resolution of the image is 0.1 arc second, an unprecedented level of detail but still nearly an order of magnitude from the camera's intended focus. Although other Hubble components are likely to reach their full design specifications after upgrades in future years, images from the Faint Object Camera may never get much better than this (see *Nature* 346, 781; 30 August 1990).



Christopher Anderson

Greenhouse effect economics

SIR—The recognition that the “social and economic issues are as important as the strictly technical” in relation to global warming (*Nature* 345, 473; 1990) is important. Economic analysis has, indeed, a number of contributions to make, of which two are particularly important. The first is to establish how much (if any) action should be initiated to limit global climate change. Second, economics can assist in identifying the most efficient (least cost) means of achieving any desired target for emission control.

John Maddox makes a valuable contribution by sketching the magnitude of the global “CO₂ business”. He argues that with fossil fuel consumption releasing approximately 25×10^{12} kg of CO₂ to the atmosphere annually, “there is no other commodity in world commerce routinely handled on such a scale”.

But this raises a central aspect of the economics of global warming. Carbon dioxide is not a commodity, and it is not traded in world commerce (except in very limited quantities). In other words, carbon emissions constitute an externality to production processes — they generate real costs to society, but costs not borne by the emitters.

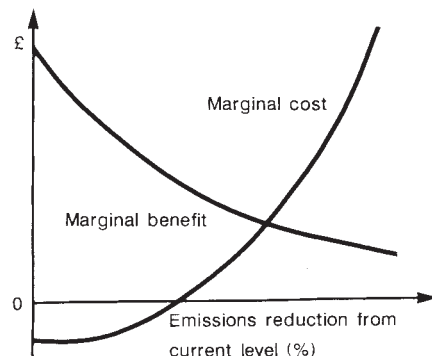
That CO₂ emissions have a zero price is the nub of the problem. Economically sensible decisions about resource use in market economies (or in planned economies where planners use ‘shadow prices’) require the presence of correct price signals. That is why economists have for a long time recommended that externalities such as greenhouse gas emissions be taxed at their true marginal social cost, so that decision makers take account of the full consequences of their decisions.

Maddox argues, correctly, that the costs of abating emissions may be very high. But his figures fail to distinguish between marginal and average (or total) control costs. This distinction is critical in considering pollution control policy. It is argued that substitution of fossil fuels by other sources would be more expensive by a factor of three, with initial capital costs at least 10 times as large. This is misleading: the marginal cost of substitution can be very low for initial fuel-switching projects, although it will tend to rise fairly sharply as the extent of substitution increases.

So we should not view the problem in terms of totals. Certainly, replacing all fossil-fuel generated electricity by other sources will be hugely expensive, but substitution at the margin may be relatively cheap. Indeed, in the short term, the cost of emission control programmes through energy conservation and efficiency schemes may actually be negative; such projects would generate positive net

present values.

We now need to focus research on quantification of the marginal cost schedule for emissions control, which is likely to have a shape like that in the figure.



Optimal policy can then be identified if we can estimate the marginal benefits function (or, the benefit society derives in the form of averted global climate change from unit reductions in emissions). Equating marginal costs and benefits identifies the best target for emissions control. Alternatively, policy makers can use such a framework to ask how severe global warming damage would have to be to justify a particular level of intervention.

This is clearly a mammoth undertaking, but is not to be avoided for that reason. Several teams of economists have been working on this problem for a number of years, and the interested reader can look at the work of William Nordhaus (Yale University) for some ‘ball park’ estimates of policy targets derived in this way. Nordhaus’ estimates that about 14 per cent of greenhouse gas emissions can be reduced at extremely low cost. Above this, marginal costs rise sharply from about \$38 per ton CO₂ at 25 per cent reduction levels to \$119 per ton for a 50 per cent reduction. Nordhaus examines three plausible marginal benefit functions (averted damage); for his low, medium and high versions, he finds the optimal targets to be greenhouse gas emission cuts of 10 per cent, 17 per cent and 50 per cent respectively.

These figures relate to optimal targets assuming the best (cheapest) option for reduction is taken at each step. The second contribution that economics can make is to guide policy makers on the appropriate policy instruments for realising least-cost control. I hope that *Nature* can contribute to this debate much as it has to our analysis of the “strictly technical” questions that it normally addresses.

ROGER PERMAN

Dept of Economics,
University of Strathclyde,
Glasgow G4 0LN, UK

1. Nordhaus, W.D. *To Slow Or Not To Slow: The Economics of the Greenhouse Effect* (in the press).

RSPCA policy

SIR—In your leading article “Escalation of animal rights war” (*Nature* 345, 647; 1990) you use a form of words bound to create ambiguity in the minds of readers.

The policy of the RSPCA, to prevent cruelty and promote kindness to animals by all lawful means, is crystal clear and always has been. Membership of the RSPCA is conditional upon adherence to that policy.

This society does not condone the breaking of the law and utterly condemns those extremists who, by their actions, place human life in danger.

D. C. SAYCE

Royal Society for the Prevention
of Cruelty to Animals,
Causeway, Horsham,
West Sussex RH12 1HG, UK

SIR—As a rational, intelligent person, I was offended by the total lack of balance in the leading article about animal rights.

Terrorist bombing is indefensible, and I abhor it, as do all my acquaintances concerned with advocacy for animals. However, the violent actions of a few do not invalidate sound ethics and reasoning.

‘Speciesism’ is a handy term for our continuing use of animals although there are no morally relevant differences to justify doing to them what we would not do to ourselves. It has been discussed by respected academic philosophers such as Peter Singer, Tom Regan, Bernard Rollin and M. A. Fox, who on reflection repudiated his own book *The Case for Animal Experimentation* in *The Scientist* (December 1986). What privileged insight entitles your author to dismiss it as “claptrap”?

The parallel drawn between anaesthesia and death evades the major issue of pain and distress inflicted on conscious animals in the laboratory. A reading of the scientific literature shows that there is no need for exaggeration: witness, for example, autotomy due to deliberately inflicted pain, described by Melzack (*Exp. Neurol.* 92, 713; 1986) and the administration of repeated inescapable shock by Anisman *et al.* (*Pharmac. Biochem. Behav.* 24, 323 & 1151; 1986). Why did the author omit equal condemnation of the grossly exaggerated claims that animal research is responsible for every medical advance (though not for any of the failures)? Consider the examples of insulin and penicillin, cited in Robert Sharpe’s book *The Cruel Deception*.

People with heart disease and ulcers (among others) may well benefit more from sound clinical investigations by physicians doing relevant and humane research than by trusting their health to results derived from animal models, as these may fail the test of extrapolation.

R. M. KESTEVEN

RRI Sydenham, Ontario KOH 2T0, Canada

Technology and the family farm

William Lesser

It is one thing for innovation in agricultural practice to be implemented thoughtlessly; quite another for it to be opposed on principle. The introduction of bovine growth hormone is a case in point.

WISCONSIN, a leading milk-producing state and, until the 1960s, home of the mix-it-yourself yellow colouring for margarine, has recently imposed a temporary ban on the use of bovine growth hormone (bGH, also called BST). Although there are unresolved matters to do with the effects of bGH on human and animal health, the main object of the ban is to protect the 'family farm'. It has been predicted that widespread introduction of the compound, which increases milk yields up to 25 per cent, will result in a 20 per cent reduction in the numbers of dairy farms, most of those going out of business being smaller farms¹. If the idea of supporting the family farm by resisting change has a historical resonance, it is indeed so, for a group of self-proclaimed neo-luddites have taken bGH as a test case for opposing advances in agriculture.

Many people would no doubt agree with the simple neo-luddite proposition that we should have a "willingness to just say no to risky innovations" (ref. 2) and technologies have typically (and rightly) been held accountable for the environmental and health problems they cause. But the neo-luddites go beyond that to examine new products and processes for their social and economic effects, and their political meanings; for them, all technologies are political, their good or evil depending on who uses them³. Their ultimate object is the development of yardsticks to measure the ill effects of technologies — yardsticks which, as with bGH, are to be applied before the release of these technologies². My view is that such analyses are not only highly unreliable, but that they overlook the hidden cost of failing to adopt change and typically respond to symptoms rather than the underlying causes. Put plainly, the neo-luddites employ sophistry in an attempt to put back the clock.

One way to appreciate the difficulty of before-the-fact analysis is to consider how hard it is to understand the effects of what has already occurred. One example is the

'green revolution', brought about by the breeding and widespread introduction in developing countries of modern crop varieties. Much has been written about how these varieties led to an increase in landless workers, to a growth in large farms and to no real improvement in the availability of food for the poorest (see, for example, the review of Lipton and Longhurst⁴). But as the same authors pointed out⁵, things are not so simple. That the nutritional status of the poorest

that the net result on labour is unclear. When mass mechanization did occur, it was encouraged by government policy. Larger farms did start to use the new varieties first, but smaller farms have subsequently adopted them successfully. The ability to purchase the necessary fertilizers, chemicals and seeds has, in general, not been a limiting factor.

It turns out that the people most disadvantaged by the green revolution are people living in areas not suitable for the new varieties, including regions with inadequate rainfall and those where rice and wheat — the dominant varieties — are not prevalent⁶. Not only are such regions denied the benefits of modern crops but higher worldwide production of small grains holds down export prices and a lack of foreign exchange restricts the import of cheaper produce. So farmers in areas not appropriate for modern crop varieties are doubly penalized.

From this brief discussion it is evident the effects of new agricultural technology are determined by the technology itself, by the particular environment in which it can (or cannot) be used, and by government policy. Because this mix is different in different countries, so are the implications. Yet the adoption of modern crop varieties by one country affects others, even those where introduction of them is not possible. Here we are left with two questions. First, who could have predicted this situation before the green revolution took off? Second, once the effects were evident, who could decide on the overall balance of benefits and costs

when some areas were favoured and others disadvantaged? Even recognizing the losses imposed on some countries, was (is) banning the use of the new varieties possible? Similarly, it is impossible to predict how bGH will perform in commercial (as opposed to experimental) herds, especially when national dairy and dairy-product trade policy are considered.

Although the neo-luddites justifiably demand that the full effects of a techno-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Nothing new under the Sun — an early luddite, disguised as a woman.

did not improve was often due to a combination of subsidies and export policies, and to rapid population growth — as a general rule, the poor, who spend the greatest proportion of their income on food (often as much as 80 per cent), should benefit most by an increase in food supply and the resulting price reduction. Labour use was affected by the options of double-cropping and intercropping made possible by shorter growing seasons, so

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Automated feed-mixing mill with self-unloading wagons for cattle in Bakersfield, California.

Joe Munroe

logy must be considered before its acceptance, they fail to take into account factors that have a great influence on the bottom line of the cost-benefit balance sheet. Never mentioned is the price of banning new products. Those such as bGH, which require registration before sale, are relatively easy to control, but it would be inordinately expensive to enforce prohibition of most mechanical innovations. Moreover, the uncertainty of a system of evaluation (which now usually operates through courts of law) reduces private investment in research. Products such as bGH cannot be properly assessed until they have passed through the development and into the experimental stage, by which time considerable sums have been invested^{8,9}. Who will make such investments in the face of uncertainty about the acceptability of the product on what have to be excessively vague social grounds?

Seemingly, the most powerful argument against bGH is that for many years the United States has produced a surplus of milk. Why increase that surplus? Yet the surplus is a result of prices set above market levels, and it cost US taxpayers about a thousand million dollars annually during the mid-1980s¹⁰. Federal costs were kept at this level by restricting imports of butter and cheese and by setting artificially high prices for milk for human consumption. These policies particularly hurt lower-income customers (who tend to have more children than the better off). The introduction of bGH can make milk cheaper; banning it has specific — if not directly visible — costs, especially

for the poor.

Support for this argument comes from an analysis of rice-threshing systems in Bangladesh, carried out by Greeley¹¹. A new mechanized system was highly profitable and relieved the drudgery of threshing by hand. The only drawback was the loss of employment for women, who traditionally were responsible for post-harvest food handling. These women were from the poorest landless class and their income contributed significantly to the welfare of their families. One way of tackling the issue of labour displacement is to abolish a new technology (if that can be done) and return to the status quo. But as Greeley points out, the real issue is not the mechanization of rice threshing but the subjugation of women in Bangladesh. As he says, in a society with a rigid gender-based way of allocating work, "The cheapest policy for the government is to intervene through programs for poor women rather than against huller mills".

With bGH, the underlying issue is the threat to the family farm, this despite the

recognition that there is nothing inherent about the product which would cause large farms to profit more per unit than small ones¹. What then is the proper (least-cost) approach? Some aspects of US policy clearly favour larger farms, notably base quota systems and, more arguably, extension/training programmes¹².

It is these programmes that should be re-evaluated. More fundamentally, size is associated with managerial ability for the simple reason that well-managed farms grow while those which are not don't. Helping small farms is a matter of helping small farmers become better managers^{13,14}. What is needed is teaching management not, as one colleague of mine has put it, teaching book-keeping.

The goal of establishing a broad social standard for evaluating agricultural technologies is unattainable. In the United States, we cannot agree at present on what constitutes an acceptable level of health risk from food products, let alone on the more involved matters of evaluating the sharing of costs and benefits across economic and national boundaries. The neoluddites are politically astute; stripped of its rhetoric, their proposal of such a process is but a publicly acceptable means of opposing innovation in general. We ought not be diverted by that approach. Any change has its losers as well as its winners, and the sharing of cost must come through the early identification of those to be penalized and the development of alternatives for them. Job retraining and management training are among the means of doing so; there are others.

Grossman identified one successful case of luddism as farm labourers who "kept threshing machines at bay for more than a generation — earning them plenty of time to accommodate the changes brought about by this technology". These days we do not have the luxury of generations to come to terms with the effects of innovation in agriculture. We must learn to live with change, and to ameliorate the difficulties of those put at a disadvantage, not to pass our time in acrimonious debate over how to resist it. □

William Lesser is in the Department of Agricultural Economics, New York State College of Agriculture and Life Sciences, Cornell University, Ithaca, New York 14853, USA.

1. Marion, B.W. & Wills, R.L. *Am. J. agr. Econ.* **72**, 326–336 (1990).
2. Grossman, D. *Utne Reader* **38**, 44–49 (1990).
3. Glendenning, C. *Utne Reader* **38**, 5–53 (1990).
4. Lipton, M. & Longhurst, R. *New Seeds and Poor People* (Unwin Hyman, London, 1989).
5. Lipton, M. & Longhurst, R. *Modern Varieties, International Agricultural Research and the Poor* World Bank Study Paper No. 2 (Washington, DC, 1985).
6. Buttel, F.H. & Barker, R. *Am. J. agr. Econ.* **67**, 1170–1175 (1985).
7. Schuh, G.E. & Tollins, H. *Costs and Benefits of Agricultural Research: The State of the Arts* World Bank Staff Working Paper No. 360 (Washington, DC, 1979).
8. Kalter, R.J. *et al.* *Cornell Univ. Dept Agric. Econ. A.E. Res.* 85–20 (December 1985).
9. Office of Technology Assessment *Technology, Public*

- Policy and the Changing Structure of American Agriculture* OTA-F-285 (1986).
10. *Agricultural Statistics 1988* (US Dept of Agriculture, 1989).
11. Greeley, M. *Postharvest Losses, Technology and Employment: The Case of Rice in Bangladesh* (Westview, Boulder, Colorado, 1989).
12. Geisler, C.C. & DuPuis, E.M. in *Biotechnology and the New Agricultural Revolution* (eds Molnar, J.J. & Kinnucan, H.) Ch. 11 (Westview, Boulder, Colorado, 1989).
13. Kinnucan, H., Hatch, V., Molnar, J.J. & Pendergrass, R. *Auburn Univ. Alabama Agric. Exp. Stn Bull.* 605 (February 1990).
14. Knutson, R.D., Yonkers, R.D. & Richardson, J.W. in *Biotechnology and the New Agricultural Revolution* (eds Molnar, J.J. & Kinnucan, H.) Ch. 6 (Westview, Boulder, Colorado, 1989).

Virtue in imperfect models

Calculation from first principles is necessary in establishing the validity of a theory, but less than rigorous intermediate models of reality retain an important function.

WHEN can an explanation of a physical phenomenon be counted as complete? It is a sufficient test that the phenomenon should be calculable from some explicit set of first principles, but there is a sense in which this is also a necessary requirement of what passes as explanation. Who, after all, would take Newton's laws seriously if it were not possible to use them for calculating the properties of any classical systems, perhaps with the help of a Cray machine?

But the role of calculation in testing a theory can easily be exaggerated. There has to be enough of it, and the sources of disagreement between calculation and measurement have to be sufficiently understood, for the community of those concerned to be convinced that the theory could be a starting point for the calculation of everything. But that does not signal that everything should be calculated. There are, after all, always errors of calculation, and one of the most persuasive criteria in assessing the validity of a theory is that successive computational improvements should yield results converging towards the measurements.

So there is a continuing need for intermediate models of reality, even rules of thumb, for predicting the properties of the real world. Numerical meteorology may have made great progress in the past few years, but weather forecasters still rightly stick with much of their ancient lore about the behaviour of the atmosphere. And despite the marvellous Cray-assisted flowering of computational chemistry in the past few years, chemists are not going to forget, for example, their familiar textbook rules about the likelihood that particular substituents will turn up at one of the three accessible positions in a monosubstituted benzene ring.

It is not simply that all computations are inaccurate (like all measurements, but sometimes more so), but that they often lack what people insist on calling "heuristic value" — at least at the beginning, they do not provide the mind with a view of reality that is easily comprehended. So it is right and proper that people should be on the look-out for intermediate models of reality.

One interesting venture of this kind is an account by Christopher A. Hunter and Jeremy K. M. Sanders, from the University of Cambridge, of how the interaction between molecules with conjugated electronic systems (as in the archetype

benzene) may be accounted for (*J. Amer. chem. Soc.* **112**, 5525; 1990). Their starting point is the habit of molecules with a planar porphyrin core (as in myoglobin, haemoglobin and chlorophyll) to form face-to-face stacks, but the phenomenon arises in many other circumstances — the stacking of purine and pyrimidine bases in DNA (and the intercalation of drug molecules therein), the stacking of successive conjugated planar arrays of carbon atoms as in graphite (and, more intriguingly, the crystal structure of conjugated molecules being canvassed as molecular electronic devices) as well as in the interaction of amino acids carrying phenyl groups in folded proteins, for example.

Computational chemists are these days well able to calculate, say, the potential energy of two benzene molecules as a function of their separation and relative orientation, but Hunter and Sanders show a trace of impatience with the results. Noting that there have been some *ab initio* calculations, they nevertheless complain that these "do not explain the basic mechanisms of π - π interactions in a way that is helpful or predictive for the practical chemist". Instead, they have developed a kind of electrostatic model of the interaction between two conjugated systems that appears to work quite well.

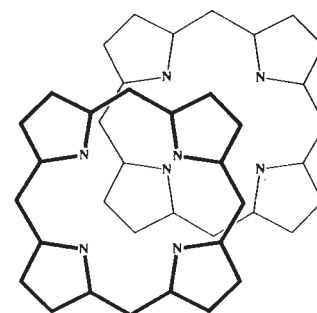
This is how the model goes. Two neighbouring carbon atoms in a conjugated system are held together by two valence bonds — one (called σ) formed by the overlap between spherically symmetrical (s) electron states, the other (called π) formed by interaction between atomic p states which, in a ring system such as benzene, have an axis of symmetry perpendicular to the plane of the molecule. The result, in benzene and other ring systems, is a two-part electron distribution concentrated along the line joining two adjacent carbon atoms (σ) and in rings tracing the outline of the ring system both above and below the plane of the molecule (π). In this simple picture, the atoms of the underlying skeleton carry positive electric charge.

So why not attempt to calculate the potential energy of two interacting conjugated ring molecules by adding together known interactions between them? Hunter and Sanders conclude that van der Waals interactions alone would lead to the maximization of overlap between the π systems in adjacent molecules, which is contradicted by observation. (In graphite,

for example, successive planes of carbon atoms are displaced by one third of the longest dimension of a six-membered carbon hexagon.)

So they represent each carbon atom in a conjugated system as a simple three-component electrostatic structure, with a central positive charge (of one unit in the simplest case) and two outlying negative charges of half a unit (for a neutral molecule) at equal distances above and below the plane. (The exact distance is derived from the measured dipole moment of benzene.) Hunter and Sanders boast that their calculations have been done with a Macintosh machine, not a Cray.

The outcome is pleasing. First, it emerges that two parallel porphyrin rings have the smallest energy when their centres are offset. (The figure shows the



calculated configuration of least energy at a separation of $3.4\text{\AA}$.) Metal ions (Zn^{2+} , Mg^{2+} or $\text{Fe}^{2+/3+}$) coordinated to the four apical nitrogens of the pyrrole rings do not change the position of the energy-minimum, but affect the energy of interaction. The authors could claim breathtaking agreement between their calculation of the offset of molecules in neighbouring planes of crystalline kekulene, the conjugated planar molecule consisting of a ring of twelve benzene rings.

Emboldened by these successes, Hunter and Sanders put forward six rules which they expect will govern the stacking properties of conjugated planar molecules. It will be intriguing to see how well they work in practice. But the simple principle that emerges from their argument is that what might have been expected to be a dominant electrostatic repulsion between the σ -electron systems is more than outbalanced by the attraction between σ electrons on one molecule and the π framework of the other. It seems unlikely that that will be overturned even by the Cray battalions.

John Maddox

Proper names for early fingers

Jonathan Cooke

A NOW eminent (and orthodox) evolutionary geneticist once claimed, in my presence, to suspect that a shark's eye and his own developed under the control of largely 'different' (non-homologous) genes, even though he still deemed them to represent equivalent episodes of morphogenesis plainly linked by common descent. It was a remark made in the early 1970s, before the methodological revolution of molecular biology and certainly pHB — pre-homeobox — but a challenge nevertheless. On page 66 of this issue Coates and Clack¹ confirm that some of the earliest undoubted (but aquatic) tetrapods sported hindlimbs and forelimbs with seven and eight digits, respectively, giving an opportunity to ask how, in the light of contemporary biology, the term homologous can best be used.

Because certain near-contemporaneous Russian fossils also reveal *bona fide* six-digit limbs², it seems that in the grade of vertebrate organization that pre-adapted its possessors for terrestrial tetrapod life, the number of digits was an evolutionarily labile character. Conceivably, it was not even well enough controlled to assure constancy between different individuals within single species. We must now accept that any pentadactyl archetype for extant tetrapod limb skeletons was not cognate with tetrapod organization. It can have become stabilized only in a subsequent lineage or lineages, the common ancestors of the extant classes. In fact, that extant fore- and hindlimb skeletons are each derived from a single ancestral form has long seemed questionable, if homologies of constituent elements are to be assigned on the bases of their relative spatial and temporal order of first cartilage condensation within the rudiments. Urodele (newt and salamander) hands and feet form in a distinctly out-of-line way³, and an anteriormost digit of five (where five exist) in other amphibian forelimbs is often surreptitiously labelled a 'prepollex' in textbooks, leaving its followers to be called one to four without further exegesis.

Yet there is little doubt that a pentadactyl ground plan lay at the base of (and has often been retained in) the longer-lasting terrestrial vertebrate radiation. The three wing fingers of the chick are not numbered two, three and four just to antagonize molecular biologists-turned-embryologists. But just how rational is this terminology? Ironically, the second, and greater, importance of Clack's and Coates' paper may be for the understanding of morphogenetic mechanisms and, by implication, the molecular basis of homology — no mean outcome from the

meticulous examination of impressions in stone. The authors support a hypothesis proposed by Shubin and Alberch³, on comparative and experimental embryological grounds, for a generative process underlying the reality of homologies between tetrapod limb elements.

Three classes of event compose limb skeletal morphogenesis: condensation of mesenchyme so that a cartilaginous mass



M.I. Coates & J.A. Clack

This hindlimb of *Ichthyostega* sports seven digits. In addition to four well-separated digits, three smaller digits lie superimposed and in a small clump, anterior to these. In the model proposed by Shubin and Alberch³, the developmental axis of the limb runs down the femur and fibula (the smaller of the two large bones distal to the femur) and along the digital arch from right to left so that the three small digits are developmentally the most distal. In later tetrapods, the axis stops at the digit thought by convention as digit 1, but the extent of axial development in very early forms such as *Ichthyostega* and *Acanthostega* seems to have been much more flexible.

will differentiate; and then both branching and segmentation events among the condensations that increase their possible numbers with overall growth in the young limb bud. A particular sequence of 'pre-axial' segmentation and 'postaxial' progressive asymmetrical branching in most extant tetrapod limbs and, it now seems, in the earliest-known ones, gives an archetypal series of elements. This is in principle of indefinite extent, but allows each element to be uniquely labelled (linguistically) according to the point in the iterative sequence where it was generated (see Figs 1 and 2 in ref. 1 and Fig. 20 in ref. 3). Simple timed perturbations of the conditions attending development (growth rate, cell number or tissue shape) can effect particular deletions, regressions or fusions in the full archetypal pattern. These resemble the modifications to pentadactyly seen in evolution, and are claimed to help explain their mechanism as well as justifying assignment of par-

ticular homologies (or proposed fusion events) among the elements.

In the scheme, most of what in descriptive anatomy is viewed as a posterior (high digit number) to anterior sequence across the limb pattern is instead seen as developmentally proximal (early in time) to distal. Digit generation is thus in principle an extendable feast, and apparently in limb rudiments of various young Devonian hopefuls went on beyond the modern stopping point that yields the pentadactyl digit one; our thumbs and big toes. Clack's and Coates' statement that the morphology revealed by the fossils would be 'impermissible' were development to have proceeded according to certain other models underestimates the adroitness of some theoreticians in assimilating data at need. Models originally proposed to illustrate the operation of constraints in evolution, I notice, usually prove pliable when faced by nature with the impermissible.

Goodwin and Trainor⁴ occupy one end of a spectrum of stances concerning the reality of homology within biological structures. For them, only the substrata of a morphogenetic episode (such as morphogens that create a concentration landscape in tissue⁵, or the dynamic viscoelastic properties of mesenchyme⁶) are homologous aspects of different ontogenies. Once the generative machinery is set going, the numbers and dispositions of repeating elements such as digits will be preserved or altered insofar as boundary conditions (for example the available amount of tissue during morphogenesis, the diffusion of morphogens or the viscoelastic constants of cells) are preserved or altered in evolution, but it makes no sense to attach identities to the parts produced across evolutionary time.

Shubin and Alberch occupy a middle ground in that they name skeletal elements meaningfully according to their lineage positions in a defined but (in evolution) open-ended scheme of condensation, branchings and segmentation. That homology is in reality even more stringent is adumbrated by the early 'nonequivalence'⁷ among the formed elements. From their inception, condensations of particular identity in the scheme may anticipate, in their growth rates, evolved special functions: the ankle bones of frogs, crocodiles and birds are examples. Thus the limb pattern ele-

1. Coates, M.I. & Clack, J.A. *Nature* **347**, 66–69 (1990).
2. Lebedev, O.A. *Dokl. Akad. Nauk. SSSR* **278**, 1407–1473 (1984).
3. Shubin, N.H. & Alberch, P. *Evol. Biol.* **20**, 319–387 (1986).
4. Goodwin, B.C. & Trainor, L. in *Development and Evolution* (eds Goodwin, B., Holder, N. & Wylie, C.) 75–98 (Cambridge University Press, 1983).
5. Turing, A.M. *Phil. Trans. R. Soc. B* **237**, 37–72 (1952).
6. Oster, G., Murray, J. & Miani, M. *J. Embryol. exp. Morph.* **89**, 93–112 (1985).
7. Lewis, J. & Wolpert, L. *J. theor. Biol.* **62**, 478 (1976).
8. Slack, J. *J. theor. Biol.* **82**, 105–140 (1980).
9. Dollé, P. et al. *Nature* **342**, 767–772 (1989).

ments are formed in a weakly but significantly structured generative sequence, and are then most probably given deep and ineffable molecular proper names, like T. S. Eliot's cats⁸. Any unique, customized programmes for the future development of elements in particular species could be identified and accessed via these names, and products of homeobox-containing gene clusters such as Hox5 (ref. 9) may be among the letters in which such names, the molecular basis of homology, are written both for the whole body and within the limb.

It is now clear that certain digits have

gone forever because the process of limb development has become truncated in evolution. But for most of us, philistine enough to accept the historically contingent nature of evolution, there is nothing specially deep about the number five. Pianists should ponder the challenge that our motor cortexes would have been set had Bach or Scarlatti sported eight deeply and ineffably named fingers per hand. □

Jonathan Cooke is in the Laboratory of Embryogenesis, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

GEOCHEMICAL CYCLES

Laughing gas from leaky pipes

Oliver C. Zafiriou

NITROUS oxide is a trace greenhouse gas that is also a major chemical control on stratospheric ozone budgets. Present in air at around 300 parts per billion, a concentration that is increasing by 0.2 per cent a year, it resides there for nearly a century. N₂O also plays important roles in marine and terrestrial ecosystems, where ubiquitous microbial processes convert a small but significant fraction of nitrogenous plant nutrients to unusable N₂O. Some of this escapes from soils and the oceans to the atmosphere. But, according to Cicerone¹, despite more than two decades of concentration-based studies, "recent assessments of [atmospheric] N₂O sources are diverging rather than converging." Oxygen and nitrogen isotope ratios could help sort out the importance of different sources. Although this was proposed and tested by Wahlen and Yoshinari², they did not analyse a broad range of natural samples owing to limitations of sensitivity. Now on page 58 of this issue³, Kim and Craig analyse a suite of N₂O samples from Pacific seawater, showing that systematic N–O isotope relationships exist. As difficult samples can now be measured with adequate precision, the approach is a step closer to application. Not surprisingly, the initial suite of results does not settle the big questions about oceanic processes or atmospheric sources, but clearly the potential exists to tackle them.

Kim and Craig have applied brute-force sampling and traditional analytical methods to vast samples obtained as a by-product of sampling for marine ³⁹Ar. During week-long stations they winched enormous 'Gerard barrels' up and down over 25 km *in toto* to obtain 40 tons of sea water. From these they extracted dissolved gases using a cryptically named "Macbeth machine". The technical advance in their work is a method for converting N and O atoms in N₂O quantitatively to N₂ and CO₂, whose ¹⁵N/¹⁴N and

¹⁸O/¹⁶O isotope ratios are then measured conventionally.

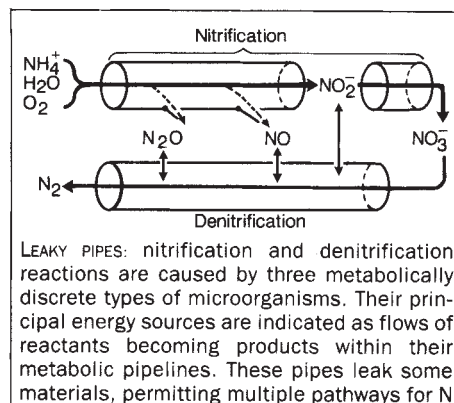
What do their data say about the mechanisms of N₂O production in the oceans? The oceanic N and O isotope ratios reported are larger than expected by previous workers, leading to conflicting interpretations of different subsets of isotope and related data^{2–4}. These cannot now be uniquely resolved. The new isotope data seem to require both the production and simultaneous consumption of N₂O or of some of its precursors, however. The box, an elaboration of the "leaky pipe" conceptual model of Mary K. Firestone, a soil scientist at Berkeley, shows that there are multiple opportunities for isotope fractionation to occur. Elucidation of which routes are significant and their isotope signatures is critical to reading the now-illegible isotopic signature of both marine and atmospheric N₂O.

Production and concurrent consumption of N₂O itself are probably partially responsible for the unexpected isotope ratios, as about 300 gigamoles a year of N₂O are produced in the sea, whereas less than 50–100 gigamoles probably outgas to the atmosphere in the same time⁵. The missing difference may well be reduced to N₂, perhaps by long-range horizontal transport to sediments where it is reduced

by denitrification⁶. Kim and Craig alternatively harmonize their data with a plausible partitioning of an intracellular nitrification intermediate. But other plausible branching pathways also occur at several stages in denitrification⁷ (see box).

The application of isotope results to unravelling atmospheric sources can be illustrated with the data of Kim and Craig's Fig. 1 on page 59. The isotope ratios in marine air are probably matched closely in water from near 700 m in the South Pacific. More importantly, appropriate mixtures of waters from a range of depths whose compositions lie along a line through the atmospheric value would also match. If the global average of waters transported to the sea surface and out-gassed resembles this mixture, the ocean source alone would match the atmospheric composition. But only a small percentage of N₂O from soil air could be added to such a mixture without violating the good fit to the marine and air isotope data — if the single reported analysis represents soil emissions. Clearly, only after obtaining representative isotope signatures and identifying their variability for all the principal sources can one derive the set of allowed source mixtures as envisaged by Wahlen and Yoshinari². How strictly sources are constrained depends on how many, how different and how isotopically variable the components are. Because the marine source⁵ is probably smaller than the total required, around 350 gigamoles per year, the air–ocean N₂O isotopic near-match may imply other similar sources.

Hence this promising approach will not quickly resolve N₂O budgets. Both the concentration and isotope approaches require globally representative samples of N₂O from source regions of great oceanic and terrestrial spatiotemporal variability. Despite thousands of relatively facile N₂O concentration analyses over about 100,000 km, oceanographers still suspect this pernicious problem. Soil emissions studies seem even more vulnerable; for example⁸, one intensively studied site less than 0.1 square degrees in area contained 16 plant communities and land use types. Because isotope data will always be rarer



and O atoms to flow into and out of the N₂O pool. Every point where streams may merge or diverge allows in principle some separation of heavy and light isotopes. Ammonia-oxidizing bacteria appear to leak N₂O, NO and possibly hydroxylamine (NH₂OH) like pressurized pipes, emitting some before they are completely metabolized. Nitrite-oxidizing bacteria are not known to leak. Denitrifying microorganisms appear in contrast to leak through figuratively porous holes which permit two-way diffusion-like passage of several intermediates. Hence the isotope signature of N₂O produced in sea water or soil may be affected by the details of processes involving nitrification, denitrification, or both.

than concentration data, representative sampling of poorly mixed reservoirs is hard to achieve or evaluate. However, combining isotope and concentration data into a three-dimensional picture may provide an invaluable tool for revealing and evaluating sampling biases and systematic errors. Surprisingly, Kim and Craig have not presented the concentration dimension of their data set.

A rigorous isotope balance may also require abandoning the assumption that only sources matter. Although N_2O probably enters the stratosphere without altering its isotope signature², its signature is certainly changed there by reactions with oxygen atoms and, in a different way, by photolysis. Hence any return flux to the troposphere of stratospheric air containing altered N_2O may enter into the isotopic book keeping. We know that some air from N_2O -depleted levels in the stratosphere returns by such mechanisms as

tropopause folding, subsidence in polar vortices, and other large-scale processes. Ronald Prinn, an atmospheric dynamicist working with global N_2O distributions, confirms that the recirculation flux cannot simply be assumed trivial (personal communication). Cicerone also emphasized¹ that some soil regions are N_2O sinks needing further evaluation; their effects on isotope ratios should be included. □

Oliver C. Zafiriou is in the Chemistry Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

1. Cicerone, R. J. *J. geophys. Res.* **99**, 18265–18272 (1989).
2. Wahlen, M. & Yoshinari, T. *Nature* **313**, 780 (1985).
3. Kim, K. & Craig, H. *Nature* **347**, 58–61 (1990).
4. Yoshida, N. *et al.* *Nature* **342**, 895–897 (1989).
5. Butler, J. H. *et al.* *J. geophys. Res.* **94**, 14865–14877 (1989).
6. Codispoti, L. A. & Christensen, J. P. *Marine Chemistry* **16**, 277–300 (1985).
7. Zafiriou, O. C. *et al.* *J. biol. Chem.* **264**, 5694 (1989).
8. Levine, J. S. *et al.* *J. geophys. Res.* **95**, 1853–1865 (1990).

NEUROBIOLOGY

A depression long awaited

Charles F. Stevens

ARTOLA, Bröcher and Singer provide, on page 69 of this issue¹, the first confirmation of a result² that has been the source of a silent controversy in neurobiology over the past year. They describe the conditions under which the strength of a synapse between two neurons is reduced with use and thereby support the earlier finding of this phenomenon, long-term depression, as the obverse face of the long-term potentiation coin.

Long-term potentiation (LTP), is widely accepted as a neuronal substrate for learning and memory. Because forgetting is such a conspicuous counterpart to learning, one might expect to find the cellular analogue of forgetting as an inverse of LTP: neuronal forgetfulness might be an enduring decrease in synaptic strength that follows the heavy use of a specific synapse.

Under what circumstances could such long-term depression (LTD) occur? LTP results when the repeated use of one of a neuron's synapses is correlated with the activity of another. Stanton and Sejnowski² thus reasoned that an anti-correlation between the activities of the synapses should favour LTD, the inverse of LTP. They designed an experiment in which they sustained a constant level of use of the synapse while switching between situations in which its activity was correlated and anticorrelated with that at other inputs. LTP occurred when synaptic activity was correlated, and this increase in synaptic strength could be reversed when activities were negatively correlated. Although Stanton and Sejnowski did not analyse the mechanism of LTD fully, they

suspected nevertheless that the voltage of a neuron is involved and noted that the phenomenon was found only when the neuron was prevented from depolarizing during stimulation of its excitatory inputs. Of course, it was already known that LTP develops through synaptic use when the neuron is sufficiently depolarized to allow influx of calcium ions through NMDA-receptor channels at its activated synapses.

This all makes great sense, but the problem remains that the precise conditions for producing LTD are not well defined, and three respected groups have been unable to find LTD using what they believed to be the conditions reported by Stanton and Sejnowski. The controversy, then, involves the very existence of LTD. Sceptics suspect that the original report involved some sort of artefact, although no one has suggested a convincing one, whereas others accept the existence of LTD and feel that the failure to confirm the original observations means that the correct procedures were not followed. The controversy has been rather muted because of a reluctance, and sometimes inability, to publish negative results.

A certain amount of confusion has been added to the controversy because, in the still-maturing field of synaptic plasticity, different phenomena have been given the same name. Several distinct types of LTP have been described which appear to have little in common at the mechanistic level. And a kind of LTD^{3–5} — interesting and important but quite different from the LTD under discussion — has been studied at particular synapses in the cerebellum and parts of the hippocampus.

Singer's group has not only found LTD as the inverse of LTP, but has also proposed a specific idea about the necessary conditions for obtaining it. According to their observations, LTD occurs when a neuron is depolarized, but not sufficiently for it to reach the threshold for calcium influx through the NMDA receptor channels. Thus neurons have a narrow window of voltages at which LTD will be produced when the synapse is used: near resting potential and below, the use of a synapse does not alter its strength; above a depolarization of perhaps 20 mV, LTP develops; but between resting potential and the LTP threshold, LTD develops and the increase of synaptic strength produced beforehand by LTP is reversed.

The experiments were done not in the hippocampus — that old standby preparation for LTP — but rather in slices from the visual cortex. And the membrane potential was regulated during synaptic use not by directly passing a current through the recording electrode, but by altering the dose of bicuculline, a drug that blocks activity at inhibitory synapses and thus permits the neuron to be more or less depolarized (according to the dose) by antagonizing the inhibitory activity that attends the stimulation used. With no bicuculline, neither LTD nor LTP was found; with small doses, LTD was observed; and with doses that blocked most of the inhibitory synaptic activity, LTP followed synaptic use. Carefully done as the new experiments are, they probably will not end the silent controversy over LTD, because some loose ends remain.

The results show that LTD can be produced if a synapse is used while the neuron's voltage is within a remarkably narrow window. If the neuron is hyperpolarized, synaptic strength is unaffected; if it is depolarized by more than about 10 mV, LTP results; in between, synaptic use causes LTD. For example, Goldman *et al.*⁶ failed to find LTD when synapses were used heavily in the presence of APV, a drug that blocks NMDA receptors. This blockage stops the flow of calcium into the synapse and LTP is not triggered. Under these conditions LTD should occur, according to the simplest interpretation of Singer's hypothesis, but it did not. Thus the controversy continues, but LTD is such an important phenomenon that we can expect a satisfactory resolution in the next year or so. □

Charles F. Stevens is at the Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.

1. Artola, A., Bröcher, S. & Singer, W. *Nature* **347**, 69–72 (1990).
2. Stanton, P. K. & Sejnowski, T. *Nature* **339**, 215 (1989).
3. Levy, W. & Steward, O. *Brain Res.* **175**, 233–245 (1979).
4. Ito, M. & Kano, M. *Neurosci. Lett.* **33**, 253–258 (1982).
5. Bradler, J. E. & Barrioneuvo, G. *Synapse* **4**, 132 (1989).
6. Goldman, R. S., Chavez-Noriega, L. E. & Stevens, C. F. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Origins of breathable air

Heinrich D. Holland

THE high oxygen content of the Earth's atmosphere sets our planet apart from all other bodies in the Solar System. It has been clear for some time that atmospheric O_2 is almost entirely the product of photosynthesis by green plants, but neither the mechanism that controls its partial pressure today nor how oxygen levels have varied during the Earth's history are well understood. Progress on the latter problem, reported at a recent meeting*, has, if anything, set back progress on the former, insofar as control mechanisms previously thought likely now seem improbable.

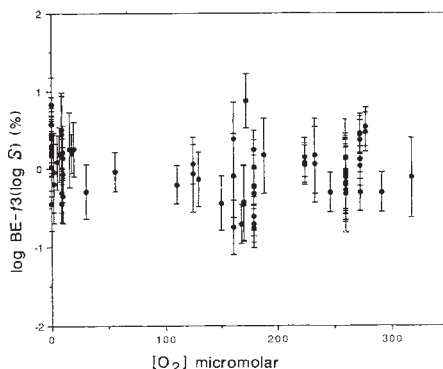
It is now generally accepted that O_2 was present in the atmosphere in vanishingly small amounts until the development of green-plant photosynthesis (J.F. Kasting, Pennsylvania State University). There is compelling evidence that more than 1,900 million years (Myr) ago, the O_2 content of the atmosphere was much lower than today, perhaps as little as 1 per cent of the present level (21 per cent). Recent studies of fossil soils indicate that the partial pressure of O_2 (p_{O_2}) rose at least fifteen-fold 1,900 Myr ago. The deposition of the enormous banded iron formations which were so important during the preceding 2 billion years ceased at that time, possibly because the rise in p_{O_2} may well have led to the oxygenation of the deeper parts of the oceans. Fe^{2+} from hydrothermal vents would then have been oxidized and precipitated rapidly on the ocean floor as oxides or hydroxides of Fe^{3+} , much as they are today. This would have reduced or eliminated the supply of iron to the shallower parts of the oceans where the banded iron formations had been deposited.

The reasons for the rapid rise in atmospheric O_2 at that time are not known. The oldest well documented fossils of eukaryotic organisms occur in rocks about 1,700 Myr old. Eukaryotes may have played a role in the rise of p_{O_2} , but they may simply have been beneficiaries of the atmospheric improvements.

The long delay — more than 1,000 million years — between the first rise of atmospheric O_2 and the development of the metazoa towards the end of the Proterozoic is still puzzling. During the past few years A.H. Knoll (Harvard University) and J.M. Hayes (University of Indiana) have shown that many limestones deposited between 900 and 600 Myr ago have anomalously high ^{13}C content, which could be due to the burial of extraordinarily large quantities of organic matter. If so, the $\delta^{13}C$ values may indicate that p_{O_2} increased very significantly

during this period; but the ^{13}C data can also be interpreted in other ways, and it is not yet certain that the evolution of the metazoa was related to a late Proterozoic increase in atmospheric O_2 .

The O_2 content of the atmosphere during the past 350 Myr is much better constrained than during the Precambrian. Currently, the most interesting limits on atmospheric O_2 during the Phanerozoic



It had been thought that the burial efficiency (BE) of organic carbon in marine sediments is influenced by the concentration of oxygen [O_2] in ocean bottom waters, so providing a feedback mechanism for controlling atmospheric oxygen levels. But this graph, compiled from 76 separate studies and corrected for differences in sedimentation rates ($f 3(\log S)$), shows it not to be the case.

are set by the presence of fusinite (charcoal) in coals deposited since the Carboniferous (W.G. Chaloner, Royal Holloway and Bedford New College; J.M. Robinson, Pennsylvania State University; N.C. Arens, Harvard University). Fusinite is an almost certain indicator of forest fires during the accumulation of plant debris. Its presence sets a lower limit of about 13 per cent on the O_2 content of the atmosphere during charcoal formation, since forest fires are unlikely to start and persist at lower O_2 concentrations.

The presence of large fossil trees in coals sets a somewhat more poorly defined upper limit on the O_2 content of the atmosphere: at O_2 concentrations in excess of 30–35 per cent forest fires would probably have been too frequent to allow the growth of large trees, and it is not clear whether the photorespiratory burden imposed by high O_2 concentrations is compatible with the record of luxuriant forest growth.

A potentially more precise O_2 barometer for the past 100 Myr may be provided by the air bubbles commonly found in amber (G.P. Landis, US Geological Survey; D. Bellis, US Department of State). For this, it needs to be demonstrated that the air in these bubbles was trapped when the amber was formed, and

that the compositional changes due to reactions with the surrounding amber can be assessed. The value of this barometer continues to be a matter of fairly heated debate. Landis reported new, but still preliminary data suggesting that the diffusivity of gases in amber is sufficiently low to prevent extensive gas transfer to and from the air bubbles. He proposed that the O_2 content of air during the Cretaceous decreased from around 35 to about 22 per cent at the base of the Tertiary (66 Myr ago), and that the percentage of O_2 in the atmosphere has remained essentially constant since then.

R. A. Berner (Yale University) described his calculations of the history of atmospheric O_2 during the past 570 Myr. These are based on estimates of the rates of burial and weathering of organic carbon and pyrite sulphur during the Phanerozoic. The approach is reasonable, since the O_2 content of the atmosphere is largely determined by these parameters, but the data base is still somewhat weak. The model suggests that the O_2 content of the atmosphere during parts of the Carboniferous and Permian (320–280 Myr ago) was about 35 per cent. The values proposed for the O_2 content of the atmosphere during the past 100 Myr disagree with those of Landis.

It is clear that p_{O_2} has varied widely during Earth history, but that it has remained within fairly narrow limits during the past 350 Myr. This period is about 100 times longer than the residence time of O_2 in the atmosphere. The partial pressure of O_2 must therefore have been controlled rather tightly during this long period. Specifically, the rate of O_2 use during weathering and due to the burning of volcanic gases must have been balanced quite precisely by the rate of O_2 production due to the burial of organic carbon and reduced sulphur and iron, largely with marine sediments.

The nature of the control mechanism is still unclear. For some time it looked as if the efficiency with which organic matter is buried with marine sediments is influenced fairly strongly by the O_2 concentration of seawater near the seawater-sediment interface, and hence somewhat indirectly by the O_2 content of the atmosphere. But this effect is currently very small, and may not exist at all (see figure). The marine geochemistry of phosphorus now seems to be the best candidate for the source of the long-sought p_{O_2} control mechanism. We need to know much more about the behaviour of PO_4^{3-} in the oceans to test this idea. Until we do, it is premature to speculate on the reasons for the major changes in p_{O_2} during the Earth's history. □

Heinrich D. Holland is in the Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.

*Atmospheric Oxygen-Variation through Geological Time
Pennsylvania State University, 16–18 July 1990.

Too close for comfort

Simon Wain-Hobson and Gerald Myers

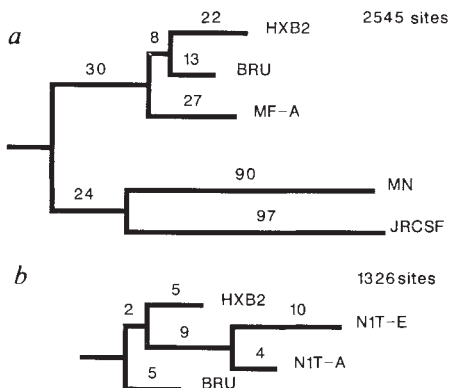
ONE of the hallmarks of the human immunodeficiency virus (HIV) is its genetic diversity. After 5 years of diligent DNA sequencing, the HIV database now contains genetic sequences from about 100 viral isolates, each of which displays a distinct genetic signature. With the exception of those obtained from a single source, or two closely related sources, pairs or groups of HIV isolates with genetic sequences differing by less than about 6 per cent are extremely rare. Against such a background it is hardly surprising that reports of HIV isolates whose sequences are strikingly similar but which supposedly originate from different patients should arouse special interest. One such case emerges from two papers appearing in the August issue of the *Journal of Virology* which provide evidence of two apparently independent, but essentially identical, HIV isolates. On the face of it these isolates would seem to call for a reevaluation of our understanding of HIV sequence heterogeneity. But do they?

First some basic facts. The genetic diversity of HIV is rooted in its replicative cycle. This involves three distinct steps: reverse transcription, DNA polymerization, and the synthesis of RNA from a DNA template (transcription). Any errors made by the polymerase enzyme during the first and third steps are not subject to proof reading, the result being pronounced sequence variability. A particular feature of the HIV genome is the existence of small duplications of 3–36 bases. These are most noticeable within the five hypervariable regions of the *env* sequence. In fact, the lability of these regions is such that genetic heterogeneity is not simply a feature of distinct HIV isolates but is also found among genomes within a single isolate. This suggests that if genetic comparisons are to be meaningful they should encompass a description of the plethora of viral genomes which, together, make up an HIV isolate.

Both E. I. Golub *et al.* (*J. Virol.* **64**, 3654–3660; 1990) and M. Stevenson *et al.* (*J. Virol.* **64**, 3792–3803; 1990) are interested in the functional heterogeneity of several different HIV-1 molecular clones derived from two viral isolates known as N1T and MF. As reported in the new papers, these isolates are from two different patients. The authors' aim is to identify small and subtle changes in the sequences of the clones which account for phenotypic differences such as infectivity and replicative efficiency.

The papers go some way towards that goal, but their particularity lies in what is not said. Thus the genetic restriction maps of the five molecular clones derived from

the MF isolate are virtually indistinguishable from those of the N1T isolate. In turn the four N1T restriction maps are strikingly similar to those of two other viruses, HTLV-IIIb and LAV-1, famous for being among the first HIV-1 isolates to be characterized in the mid-1980s. The clustering of MF-A, N1T-E, N1T-A, BRU and HXB2 is evident in a phylogenetic tree



Tree analysis produced by Swofford's PAUP program and based on (a) *env* sequences, and (b) *vif* plus LTR (long terminal repeat) sequences. Each analysis included the same set of 9 other North American sequences. Only the nearest neighbours MN and JRCSF are shown. BRU is from LAV-1 whereas HXB2 is from HTLV-IIIb. N1T-A/E and MF-A are equidistant from both BRU and HXB2.

analysis of the DNA sequences reported in the papers (see figure).

The similarity of these sequences runs counter to expectations. Indeed so much so, that when one realizes that both the N1T and MF isolates were derived in the same facility, N1T first and then MF, and that authors of both papers cosigned an earlier paper on the biological properties of the N1T isolate, the notion that MF could in fact be N1T, rather than an independent isolate, begins to take hold. By the same token, could the similarities between N1T and the prototypical HIV-1s — HTLV-IIIb and LAV-1 — reflect an extension of the chain? Can we be absolutely sure that N1T did not originate from LAV-1?

Although none of this detracts from the validity of the papers — both groups worked with molecular clones so there was no need for them to describe the inherent sequence variability of their isolates — it does deserve comment. But another paper should also be noted. Earlier this year, T. McNearney *et al.* (*Proc. natn. Acad. Sci. U.S.A.* **87**, 1917–1921; 1990) reported sequence data derived from seven apparently independent HIV isolates from a virus donor and two recipients. What is astonishing is that these authors discovered only four variable

nucleotides within 13 entire *env* sequences. And of course all sequences were co-linear throughout their hypervariable regions. There is nothing like this in the entire HIV sequence database, which now contains more than 1,000 sequences. A number of other features, notably that all the *env* products are functionally defective, make these data very hard to assimilate — they simply buck the trend.

So, either we have some awkward observations that threaten to overturn the established view of HIV genetic diversity, or else these sequences are not quite what they appear to be. At this stage, the possibility of viral contamination should not be ruled out. Contamination of cells is well known to all cell biologists. Thus, SIV contamination has already been described, as has contamination of genomic DNA by phages and plasmids. Added to this is the threat posed by contaminant amplification through the ubiquitous use of the polymerase chain reaction (PCR). One of us (S.W.-H.) has encountered more than a dozen cases of PCR contamination due to laboratory plasmids and M13 phages.

No one can say for sure whether the sequences of the clones derived from MF are in fact those of clones inadvertently derived from an N1T contaminant. But this possibility remains, given that materials were exchanged between the two groups of authors.

Are such issues important? The purity and authenticity of biological reagents and data are always crucial, but especially so when the identity of an HIV-1 sequence is used as an indication of where the virus came from. The widely assumed genetic diversity of HIV isolates is currently being turned into the basis of a method for establishing unusual or rare transmission routes, a development that raises the stakes considerably. A recent investigation into the possible transmission of HIV from a dentist to a patient during an invasive oral procedure, for example, revolved entirely on the similarity of two HIV-1 sequences (*Morbidity and Mortality Weekly Report* **39**, 489–493; 1990).

We believe that as much attention and rigour must be used in the molecular characterization of HIV-1 isolates as is necessary in human DNA fingerprinting. To establish the identity of an HIV isolate, we suggest that a population of sequences should be sought. If a viral population is not observed, or if any two are uncannily similar, then the onus should be on the sequencers to establish the origin of the viral isolate and lay to rest any fears about contamination. Given the current state-of-the-art technology, those of us in AIDS research should not settle for less. □

Simon Wain-Hobson is at the Institut Pasteur, 28, Rue Du Dr Roux, 75724 Paris Cedex 15, France. Gerald Myers is at the Los Alamos National Laboratory, New Mexico 87545, USA.

Old and young mantle roots

Paul F. Hoffman

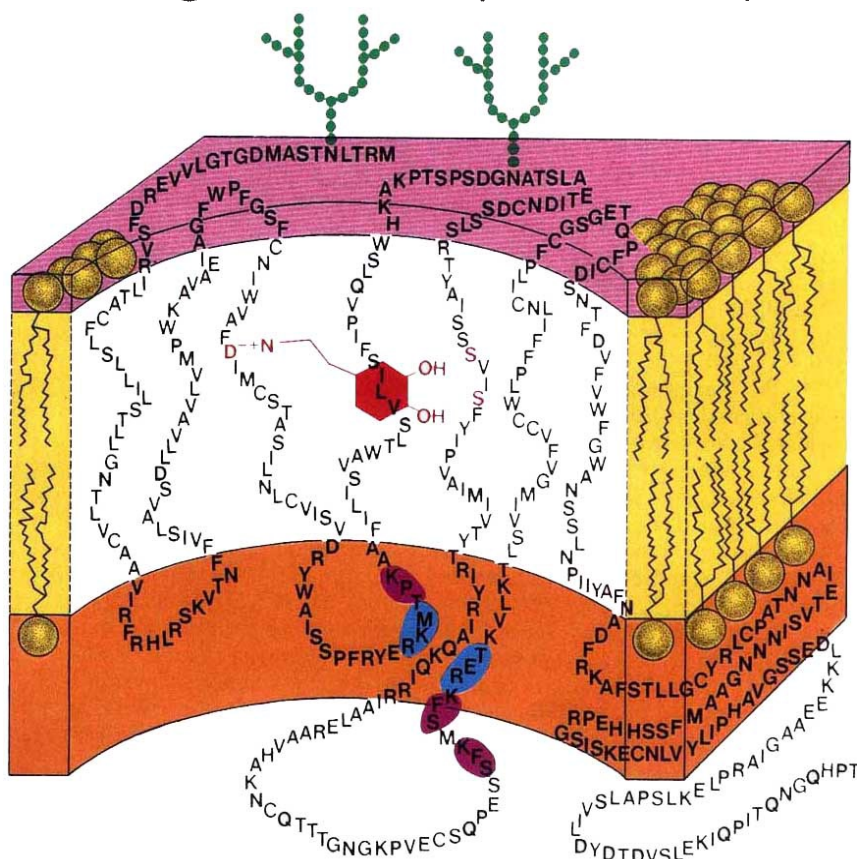
It is generally believed that the Earth has cooled over the past 4 billion (10^9) years owing to the slow dissipation of the heat originally generated during the planet's growth and the formation of its core, and owing to the decline in internal radioactive heating. But in the absence of empirical constraints, we can specify *a priori* neither the efficiency of heat transfer from the deep interior to the surface (which is related to the form of mantle convection and existence of thermal boundary layers) nor the content and distribution of radionuclides. One way of deducing thermal regimes is by estimating the flexural strength and equivalently the effective elastic thickness of stable continental lithosphere. This thickness corresponds to the temperature-dependent depth of the transition zone above which rocks behave elastically and transmit heat conductively, and below which they deform by power-law creep, fluid-like on geological timescales. By analysing lateral variations in the thickness of ancient sedimentary layers resulting from lithospheric flexure, Grotzinger and Royden demonstrate on page 64 of this issue¹ that the lithosphere in a part of the Canadian shield that stabilized at least 2.5 billion years ago (2.5 Gyr) had an effective elastic thickness at 1.9 Gyr that was almost an order of magnitude less than its estimated² present thickness, 100 ± 25 km. This result implies that the thermal gradient has decreased by a factor of 2–4 since, which is consistent with estimates of secular cooling of the Earth. But it is at odds with other indications which have fostered the paradoxical view^{3–6} that, despite a presumed high heat flux through the mantle, thick cold roots existed beneath Archaean cratons by 3.2–2.7 Gyr.

If the Earth's mantle were uniform in composition, the limiting thickness of the lithosphere would be the same everywhere. In fact, Archaean cratons (continental areas which have not been significantly stretched or shortened since 2.5 Gyr) appear to have cold mantle roots that extend twice as deep (to about 200 km) as the lithosphere beneath younger continental crust and thermally mature ocean basins. This conclusion is supported by the correlation of Archaean cratons with areas of low surface heat flow⁷, high shear-wave velocities^{8,9} and high lithospheric flexural rigidities². As emphasized by Jordan¹⁰, the existence of cold mantle roots (tectosphere) implies that Archaean cratons are underlain by anomalous mantle material that is both refractory and of low intrinsic density (to offset its lower temperature). These requirements are met by mantle material that has been

depleted in major elements by partial melting¹¹, an inference strongly supported by the compositions of rock fragments (xenoliths) brought up in volcanic pipes from beneath Archaean cratons¹¹. Thus, the depleted roots represent distinct mantle reservoirs that impart characteristic mechanical properties to cratonic plates; they demand genetic explanation.

A key question is whether the mantle roots accumulated progressively (so being thickest beneath the oldest plates) or were a product of unique conditions existing in the Archaean. The latter view has been ascendant. First, it was argued that the estimated Archaean heat flux would have resulted in continuous melting of the continental crust, contrary to geological evidence, were it not protected by refractory mantle roots³. In fact the pressure–temperature conditions implied by equilibrium mineral assemblages require no more heat flux to the base of continental

Cloning of human dopamine receptor



THREE papers in this issue^{1–3} report the cloning and expression of the gene that encodes the human D_1 dopamine receptor. Both the D_1 and D_2 receptors are targets of drug therapy in disorders such as Parkinson's disease and schizophrenia. The picture shows¹ that the protein comprises 446 amino-acid residues (single-letter code) arranged in 7 membrane-spanning segments. Dopamine (red) is thought to bind to aspartate (red letter D) and two serine residues (red letter S). The phospholipid bilayer of the cell membrane (yellow) is bounded by the extracellular (pink) and cytoplasmic (orange) aspects. Polysaccharides (green) are

attached to glycosylation sites (asparagine residues, letter N). Purple blobs highlight residues thought to be substrates for protein kinase C residues involved with protein kinase A are blue.

The end result of dopamine binding to the D_1 receptor is the activation of adenylyl cyclase, a process mediated by G proteins. The D_2 receptor, in contrast, inhibits adenylyl cyclase. The crucial difference may come from the third cytoplasmic loop and the carboxyl-terminal tail (bottom right). The D_2 receptor has a long loop and a short tail, whereas the D_1 receptor has a short loop and a long tail. This allies the D_1 receptor with the β_2 -adrenergic receptor, which also works by virtue of its interaction with G proteins. □

1. Dearry, A. et al. *Nature* **347**, 72–76 (1990).
2. Zhou, Q.-Y. et al. *Nature* **347**, 76–80 (1990).
3. Sunahara, R. K. et al. *Nature* **347**, 80–83 (1990).

Philip Seeman

crust in the Archaean than occurs now¹.

The most persuasive discovery was that diamonds occurring in 90-million-year-old volcanic pipes in the Kaapvaal craton of South Africa contain microscopic inclusions, presumably cogenetic, of sub-calcic garnet (absent from diamonds outside the craton) which have isotope ages of 3.3–3.2 Gyr, implying⁵ the persistence of a mantle root fixed to the craton extending to the depth of diamond stability (below about 150 km) since at least 3.2 Gyr. Finally, velocities of seismic waves travelling vertically through the Superior craton in Canada have an azimuthal anisotropy implying that the mantle is strained down to around 200 km deep; this reflects the bulk flattening plane in the overlying crust resulting from deformation at 2.7 Gyr (ref. 6).

Were the depleted mantle roots formed so long ago, a plausible explanation would be that Archaean cratons were mechanically underplated by buoyantly subducting slabs of oceanic lithosphere¹². Buoyant subduction was presumably common in the Archaean because the excess heat loss associated with sea-floor spreading would have resulted in many young, hot plates having thick crusts and strongly depleted mantle residues¹³.

Now, the flexural analysis by Grotzinger and Royden¹ contradicts the inference that the roots formed in the Archaean. The sedimentary rocks they studied were deposited near the margin of the Slave craton which collided with and was partly overridden by another craton to the south-east. The consequent north-west-directed thrusting deflected the Slave craton, creating a foreland basin that deepened to the southeast¹⁴. Subsequent warping and erosion has exposed a continuous cross-section of the basin for nearly 350 km across the depositional strike and this, combined with superb outcrop and lack of penetrative deformation, makes the area a sedimentological paradise. The authors carried out a comprehensive palaeo-environmental analysis, including the mapping out of isochronous sequence boundaries and large-scale features such

as prograding clinoforms and stacked forebulge unconformities normally seen only in high-resolution seismic reflection profiles.

Their result is that the craton then was considerably thinner than it is now, in direct disagreement with the arguments outlined above. So, where do we go from here? As the tectonic history of the Slave craton may well have differed from that of other Archaean cratons, it is crucial that similar flexural analyses be carried out on those cratons where evidence exists for mantle roots of Archaean age. The Witwatersrand (2.7 Gyr) and Transvaal (2.5 Gyr) basins on the Kaapvaal craton and the Huronian basin (2.4 Gyr) on the Superior craton may be suitable for such studies, which would test the congruency of the different means of characterizing

cratonic lithosphere in the geologic past. In the meantime, we are left to speculate on the means by which the effective elastic thickness of the Slave craton increased by almost an order of magnitude between 1.9 Gyr and the present. Perhaps it is time to dust off some previous notion, such as progressive buoyant accretion onto the craton base of depleted mantle bodies derived from slabs of oceanic lithosphere heated after their descent in subduction zones¹⁵. In that picture, the thickness of depleted mantle should increase with the age of the plate, exponentially so if the rate of accumulation has declined since the Archaean. □

Paul F. Hoffman is in the Continental Geoscience Division, Geological Survey of Canada, Ottawa, Ontario K1A 0E8, Canada.

ATMOSPHERIC DYNAMICS

Ozone depletion in the Arctic

Alan Plumb

To what extent, if at all, is ozone being depleted in the Arctic? This has been a burning question ever since the discovery¹ of the Antarctic ozone hole and its identification as a consequence of chemical destruction of ozone, primarily via chlorine-catalysed reactions. The answer, according to Proffitt *et al.* on page 31 of this issue², is 'a lot'. In an analysis of measurements³ taken during the Airborne Arctic Stratospheric Expedition (AASE) in January and early February 1989, the authors claim that 50 per cent or more of the ozone content of high-latitude air parcels between 17 and 20 km altitude was destroyed during the winter.

There is no dramatic 'hole' in the Arctic winter or spring to compare with that in the Antarctic (despite some misleading references to an Arctic hole); nevertheless, an apparent long-term downward trend in ozone concentrations in the Northern Hemisphere⁴ may indicate losses there. Although temperatures in the large-scale vortex circulation over the Arctic in winter never fall as low as they do in the Antarctic vortex, and although the Arctic vortex is relatively short lived, lower-stratospheric polar temperatures in the Arctic winter frequently fall low enough to permit the formation of the all-important polar stratospheric clouds (PSCs). These clouds are an essential component of the process by which chlorine is released from its stable reservoirs into active forms, thus priming the vortex for ozone destruction.

Such considerations were the motivation for the AASE. As in the preceding Antarctic mission of September and October 1987, the chemical composition of the Arctic lower stratosphere was found to be highly anomalous; in fact, late in the

mission, peak concentrations of the key radical ClO approached those seen previously in the Antarctic⁵. In recently published calculations^{6–8} based on those observations and on current knowledge of the relevant reaction rates, ozone loss during the mission was estimated to be around 12 per cent near 20 km altitude (with greatest loss rates late in the mission), a figure consistent with *in situ* ozone measurements, once allowance is made for local resupply by descending air in the vortex⁹.

Proffitt *et al.* argue that the depletion of ozone at these altitudes over the winter as a whole was considerably greater than this. Over the period of the actual mission, their calculated losses are in fact consistent with those noted above. What is most remarkable is the author's claim that about two-thirds of the total loss occurred before the mission began. They argue that this could have occurred earlier in the winter at higher altitudes, where temperatures were cold enough to permit the formation of PSCs, with the depleted air then being advected down by the prevailing polar subsidence.

The argument Proffitt *et al.* use begins with the observation that, outside the polar vortex, lower-stratospheric measurements reveal a simple correlation between local mixing ratios of ozone and nitrous oxide. N₂O has a long lifetime and so is conserved by air parcels. In the absence of anomalous chemistry, the same is true of ozone in this region; therefore air parcels moving around will conserve both mixing ratios, thus preserving the relationship between them. Inside the vortex, the O₃–N₂O relation breaks down.

How one interprets this difference between the air within and outside the vortex depends on the assumptions one

1. Grotzinger, J. & Royden, L. *Nature* **347**, 64–66 (1990).
2. Bechtel, T.D., Forsyth, D.W., Sharpton, V.L. & Grieve, R.A.F. *Nature* **343**, 636–638 (1990).
3. Davies, J.F. *Earth planet. Sci. Lett.* **44**, 231–238 (1979).
4. England, P. & Bickle, M. *J. Geol.* **92**, 353–367 (1984).
5. Richardson, S.H., Gurney, J.J., Ertank, A.J. & Harris, J.W. *Nature* **310**, 198–202 (1984).
6. Silver, P.G. & Chan, W.W. *Nature* **335**, 34–39 (1988).
7. Ballard, S. & Pollack, H.N. *Earth planet. Sci. Lett.* **85**, 253–264 (1987).
8. Lerner-Lam, A.L. & Jordan, T.H. *J. geophys. Res.* **92**, 14007–14026 (1987).
9. Grand, S.P. *J. geophys. Res.* **92**, 14065–14090 (1987).
10. Jordan, T.H. *J. Petrology, Spec. Lithosphere Issue*, 11–38 (1988).
11. Boyd, F.R. *Earth planet. Sci. Lett.* **96**, 15–26 (1989).
12. Helmstaedt, H. & Schulze, D.J. in *Kimberlites and Related Rocks* (ed. Ross, J.) Vol. 1, 358–368 (Blackwell, Carlton, Australia, 1989).
13. Bickle, M.P. *Earth planet. Sci. Lett.* **80**, 314–324 (1986).
14. Tirrul, R. & Grotzinger, J.P. *Tectonics* (in the press).
15. Oxburgh, E.R. & Parmentier, E.M. *Phil. Trans. R. Soc. A288*, 415–429 (1978).

makes about the meteorology of the polar stratosphere. Numerical studies of the dynamics of such vortices have led to the concept of a containment vessel¹⁰, meaning that the vortex boundary is impermeable to inflowing air. This picture is apparently consistent with analyses of the evolution of tracers within the Antarctic vortex¹¹ and with observations^{12,13} of strong gradients in tracer mixing ratios at the edge of the Arctic vortex during the AASE. In this view, the air parcels inside and outside the vortex comprise independent populations with different histories, in which case the differences in the O₃-N₂O relationship are no surprise.

As Proffitt *et al.* point out, however, this picture cannot be entirely correct. It is known that, during a normal winter, air subsides in the polar regions (thus, for example, maintaining the polar temperatures well above radiative equilibrium values) and this subsidence must be balanced by inflow. The issue is whether this inflow is large enough to have a significant impact on trace species budgets within the vortex. Proffitt *et al.* maintain that it is. They regard the Arctic vortex as a flowing processor in which air flows (apparently rather freely) across the vortex edge from the outside to the inner regions. Then, to rationalize the different O₃-N₂O relationships in these regions, it must be concluded that ozone is not conserved in the process. By comparing the ozone content of air with the same N₂O mixing ratio inside and outside the vortex, they conclude that air parcels that originate outside the vortex near 20 km altitude with ozone mixing ratios of 3 parts per million (volume) lose more than half that by the time they arrive in the interior. If this is true, then ozone losses in the Arctic in winter are not very much less than in the Antarctic spring. The effects on the column-integrated ozone are much less dramatic than in the Antarctic, however, because (in this hypothesis) polar ozone is resupplied by the poleward circulation about as fast as it is being destroyed.

This work emphasizes the need for a detailed, quantitative model of the dynamical meteorology of the polar vortices if we are to reach a satisfactory understanding of ozone depletion. Routine stratospheric meteorological observations alone are neither sufficiently accurate nor comprehensive enough to give us the detail needed and it is necessary to rely on indirect means to construct a full picture. Nevertheless, it is difficult to reconcile the authors' premise of substantial inflow with what we do know. Subsidence rates during AASE within the vortex, deduced from radiative calculations¹⁴ and from observations⁹ of N₂O, were about 0.5 mm s⁻¹; mass continuity then implies inflow across the vortex edge in the altitudes of interest of no more than about 0.07 m s⁻¹ (less than this if the inflow

maximizes higher up) or about 200 km per month. At this rate, the latitudinal displacements that Proffitt *et al.* appear to assume (up to 1,500 km) could not be attained during the winter.

Moreover, air moving systematically poleward will, like water in a bathtub (and for the same reason — conservation of circulation) accelerate eastwards; a net displacement of 1,500 km at these latitudes would thus result in a net acceleration of about 200 m s⁻¹. But observed winds indicate that, on the contrary, air parcels (if, indeed, they move in this way) must in fact decelerate in the process. Such movement could be aided by nonaxisymmetric wave motions, whose effects would decelerate the flow. Given estimates of this deceleration under normal winter conditions of 1–2 m s⁻¹ per day (and there are reasons to believe that it was probably smaller than this in January 1989, especially inside the vortex) it is just conceivable that this kind of displacement could have occurred over the entire winter, but not by the beginning of January. Finally, significant flow across the vortex edge would be difficult to reconcile with the observed maintenance

of strong gradients of tracers and of potential vorticity (a dynamical tracer)¹⁵ there. For these reasons many atmospheric scientists will be sceptical of the conclusions offered by Proffitt *et al.*, at least until a dynamically consistent model of the polar vortex as a flowing processor can be described. □

Alan Plumb is in the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

1. Farman, J.C. *et al.* *Nature* **315**, 207–210 (1985).
2. Proffitt, M.H. *et al.* *Nature* **347**, 31–36 (1990).
3. *Geophys. Res. Lett.* **17**, No 4 (1990).
4. *Scientific Assessment of Stratospheric Ozone: 1989* (World Meteorological Organization, Geneva, 1990).
5. Brune, W.H., Toohey, D.W., Anderson, J.G. & Chan, K.R. *Geophys. Res. Lett.* **17**, 505–508 (1990).
6. Isaksen, I.S.A. *et al.* *Geophys. Res. Lett.* **17**, 557–560 (1990).
7. McKenna, D.S. *et al.* *Geophys. Res. Lett.* **17**, 553–556 (1990).
8. Salawitch, R.J. *et al.* *Geophys. Res. Lett.* **17**, 561–564 (1990).
9. Schoeberl, M.R. *et al.* *Geophys. Res. Lett.* **17**, 469–472 (1990).
10. Jukes, M. & McIntyre, M. *Nature* **13**, 590–595 (1987).
11. Hartmann, D.L. *et al.* *J. geophys. Res.* **94**, 11625–11640 (1989).
12. Lait, L.R. *et al.* *Geophys. Res. Lett.* **17**, 521 (1990).
13. Lowenstein, M. *et al.* *Geophys. Res. Lett.* **17**, 477–480 (1990).
14. Rosenfeld, J.E. *et al.* *Geophys. Res. Lett.* **17**, 345 (1990).

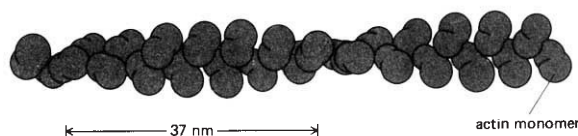
PROTEIN STRUCTURE

The changing shape of actin

David J. De Rosier

THE muscles with which we move, the delicate stereocilia we need to hear and the microvilli that absorb nutrients from our last meal share a common structural thread — the actin filament. Every cell in our bodies possesses actin filaments (see figure) built into a variety of subcellular machinery. W. Kabsch, K. C. Holmes and colleagues now provide the first view of the atomic structure of this essential ele-

ment (ATP or ADP) and a calcium ion. The DNase molecule sits on top of the cleft. The authors further divide each domain into two subdomains, so that in the actin molecule, the four subdomains lie roughly at the corners of a square with the cleft splitting the edge. In their second paper, the authors produce an atomic model for the filament based on the atomic structure of the subunit and on X-ray diffraction



The actin filament appears to consist of two helical, intertwined strands of actin monomers. These monomers aggregate spontaneously to form filaments, unless blocked by DNase or profilin: this inhibition is necessary for the crystallization of actin monomers for X-ray diffraction studies. Picture from *Molecular Biology of the Cell* 2nd edn, (Alberts, B. *et al.*) (New York, 1989).

ment in the cytoskeleton. On pages 37 and 44 of this issue, two papers present the atomic structure of the actin subunit (monomer) complexed with the enzyme DNase I as well as an atomic model for the structure of the filament, a helical polymer of actin subunits.

As seen in the 2.8-Å map of the actin:DNase I complex in the first paper, the actin subunit consists of a large and a small domain between which is a cleft containing a bound molecule of adenine nucle-

data from gels of oriented actin filaments. Lacking the phases of the X-ray reflections from the gels, they could not generate an electron-density map as they did for the actin:DNase crystals. Instead, they generated models of filaments by arranging the subunit in different ways with the appropriate helical symmetry. Of course, each possible orientation of the subunit yielded a different model. Holmes and colleagues sampled the full range of possibilities, judging their success from the agreement of the diffraction pattern calculated from each model with that observed from the oriented gel.

In their resulting filament model, the small domain lies farther from the filament axis than the large domain, thus generating a filament 90–95 Å wide. The line between the two domains is nearly

perpendicular to and slightly skewed with respect to the filament axis. Using data from the literature, the authors were able to identify the pointed end of the filament which determines the direction in which actin is translocated by myosin and locate the myosin binding site on the large domain of the actin monomer. In the filament, each actin subunit interacts with four of its neighbours and the contacts seem to involve all four subdomains. The most extensive interactions between subunits are made axially (along each of the twin actin strands rather than between them). The model explains many of the results from chemical and spectroscopic measurements — for example the accessibility of lysine residues — thus lending further credibility to its general correctness. Using data from the literature, the authors were able to identify the pointed end of the filament which determines the direction in which actin is translocated by myosin.

In these two papers, though, the authors refer to the atomic structure of the actin:DNase I complex as well as to the atomic model of the actin filament. Are these not both atomic models whose correctness can be judged by their fit to X-ray data? The distinction here lies in the certainty with which the atoms can be located. With a 2.8-Å-resolution map obtained by isomorphous replacement, the positions of atoms can be determined to about 0.5 Å. In producing a model of a filament, however, the authors assumed that the subunit conformation as seen in the complex remains unchanged in the filament. This is the proper first approximation to adopt, and indeed, the model agrees reasonably well with the X-ray data and seems to explain many other observations. In the next step, the authors have begun to refine the model by allowing the subdomains to move relative to one another. In so doing one of the subdomains has shifted by 15 degrees. Thus, as refinement proceeds, significant changes in atomic positions could occur that would affect the details of the interatomic interactions, although the broader features of the filament model will be unaffected.

The atomic model, although the most detailed, is not the only model of the actin filament. There are various maps and models derived from electron micrographs of either filaments on their own or with accessory proteins, such as myosin. All these views reveal differences that have been the source of much debate. Will the X-ray model now take precedence? The important caution is that observed differences between models may result from real differences in structure: for example, the presence of myosin or its S1 fragment might alter the structure of actin significantly.

Moreover, actin not only binds ATP but

also catalyses its hydrolysis to ADP. The role of this reaction is intriguing, because the energy released by the reaction could be used to alter the conformation of actin. Indeed, P. A. Janmey and colleagues find that ADP-actin filaments may adopt two conformational states. On page 95 of this issue, these authors report that ADP-actin filaments made by hydrolysis of ATP-actin filaments are rigid whereas those filaments made by polymerizing ADP-actin subunits are flexible. They suggest that in the hydrolysis of ATP to ADP, energy is trapped and stored in the filaments. Such energy would be available for mechanical work in the cell.

The notion that actin filaments might perform work in a cell in the absence of myosin seems heretical given the exciting recent results on myosin motors (see J. A. Spudis, *Cell Regulation* 1, 1–11; 1989). In the formation of the acrosomal process in marine sperm cells, however, the polymerization of actin or the uncoiling of an actin bundle can generate motion without any evidence of help from myosin, although other proteins are involved. The movement is quite dramatic: a 60–90-micrometre-long cell extension shoots out at about 10 micrometres per second. In the test tube, even in the absence of other proteins, the actin filament can vary its structure. These changes have been variously explained as angular disorder, lateral slippage of the two substrands and a change in flexibility. It is unclear what causes these changes or indeed whether they are all manifestations of the same phenomenon.

Holmes, Kabsch and colleagues have compared the structure of the ADP-actin-DNase complex with that of the ATP-actin complex but see no differences except, of course, for the additional phosphate. Perhaps this is not unexpected, because the presence of DNase inhibits the ability of the actin subunit to exchange bound nucleotide with nucleotide in the medium. The authors point out that the DNase molecule sits on top of the nucleotide cleft, where it can lock the cleft in one conformation and might also prevent any changes resulting from nucleotide exchange. By contrast, actin has an enhanced ability to exchange bound nucleotide when complexed with the protein profilin. Profilin sits at the opposite end of the actin subunit, at the bottom of the cleft, and could therefore hold the cleft in a different, open conformation. Fortunately, elucidation of the structure of the actin-profilin complex by C. E. Schutt and his colleagues at Princeton is nearing completion: we might soon be treated to a view of the actin subunit in one of its other conformations. □

David J. De Rosier is at the Rosentiel Center, Brandeis University, Waltham, Massachusetts 02254-9110, USA.

Water at bay

How does the non-stick coating stick to the frying pan? Not very well, is the usual answer; a firmly adhering non-stick coating is almost a technical contradiction. But Daedalus has a way out. He is anxious to exploit graphite fluoride, the non-stickiest, most unwettable material of all; and plans to form it as an integral chemical part of the surface to be coated.

Most organic materials can be decomposed to graphite by simple heating: this is how carbon fibre is made. Ultra-rapid laser-pulse irradiation can raise the surface of a solid to a high temperature while leaving the interior quite cold. So DREADCO's physicists are scanning lasers across various solid polymeric samples, decomposing their surfaces to a thin layer of graphite still chemically bonded to the material beneath. Subsequent treatment with fluorine should then give the ultimate non-stick, non-wet, firmly attached graphite fluoride surface coating.

DREADCO's first 'Grafluorinated' products will be plastic windows and lenses. Their novel surface 'bloom' will reduce reflections, but will also be uniquely water repellent. Goggles and spectacles with the new lenses will never mist up. Droplets condensing on them will be unable to adhere and will simply slide down. Raindrops, instead of giving the curious splotched and haloed outlook known to all spectacle wearers, will just bounce off. Windscreens made of the new material will need no wipers, and windows made of it will never need cleaning. Unable to adhere, dirt will just blow away; even frost will slide helplessly off.

In the same way, a Grafluorinated paint surface would never hold the dirt, and would be so water repellent that the metal underneath could never rust. The implications for the car industry are truly awesome, and those for marine engineering even more so. A true non-stick non-wet ship's paint should repel barnacles and other fouling organisms totally, giving wonderfully low drag. Even better, imagine a hairy carbon-fibre coating, flock-sprayed onto the ship's surface and fluorinated *in situ*. The unwettable fibres would retain the air between them as a thin layer, and the boat would hardly touch the water at all. Like the water-vole (whose water-repellent coat works a similar trick) it would skim along on a thin layer of lubricating air, saving a vast amount of fuel. Similar non-wet internal coatings for pipes and hoses should transform the pumping efficiency of the water industry, the fire-fighting services and chemical engineering generally. And Grafluorinated clothes would repel water perfectly even if you fell in, while being so dirt proof that you could simply shake them clean.

David Jones

Climate change and methane

SIR—Among the many explanations of the rapid climate changes that occurred at the end of the last glaciation is the suggestion that warming was initially driven by methane emitted from northern gas fields and gas hydrates, during a time of rising insolation^{1,2}. This explanation is compatible with the recent reports^{3,7}, that warming took place in two abrupt pulses, 13,000 and 10,500 years ago.

The 'methane-led' account of deglaciation invokes the strong positive feedback that would follow evasion of gas stored at high latitude during glaciation. In the model, increasing insolation destabilizes methane that accumulated during glaciation. The methane became trapped in hydrate when upward seepage was blocked by low temperatures, in areas such as the giant West Siberian fields, north-west North America, or the southern North Sea. Methane is incrementally a much more effective greenhouse gas than CO₂, and the rupture of one or two large gas pools (for example, 10^{13–14} g CH₄), or emission from an area of permafrost, could have initiated a thermal runaway, liberating more gas (possibly 10¹⁵ g CH₄ per yr) and rapidly raising planetary temperatures. The liberated gas would have been converted to CO₂ in the air, and the atmospheric warming would have induced oceanic warming and discharge of further CO₂, thus replenishing the inventories of carbon both on land and in the air.

The positive feedback would have decayed as the thermal pulse penetrated into the strata of the gasfields, according to the square root of time elapsed. Thus the initial methane-pumped runaway would be chaotic and short-lived — a few centuries — followed by a sharp decline in atmospheric CH₄, which has a lifetime of a few years, together with a slower decline in atmospheric CO₂ as the land biosphere captured carbon and the oceans again cooled. A second warming pulse might then follow, induced by the destabilization of hydrate in the extensive regions inundated after the first pulse by marine transgression and ice-dammed lakes, and aided by the further increase in insolation.

The hypothesis has the merits that it explains the speed and linkage of the rises in atmospheric CH₄ and CO₂ (refs 4, 5), and the brief standstill in the ¹³C record⁶. It demands, in particular, a sharp initial thermal pulse. The initial melting of the ice sheets around 13,000 yr BP was very abrupt³, with a sea-level rise of 24 m in less than 1,000 yr. This pulse occurred at the same time as the ¹³C age plateau, followed by a much slower change in sea level and then a second melting pulse around 10,500 yr BP. Insolation in the Northern Hemisphere was at a maximum around 11,000

yr BP and varied smoothly: there is no astronomical explanation for the sharply defined melting pulses. The Younger Dryas cooling episode which has a much stronger signal in the Northern Hemisphere is best explained as the record of a transient⁸ response to a decline in methane forcing.

The sharpness of the pulses, especially the first, is consistent with runaway methane-driven warming. There are, of course, many other possible explanations of the melt pulses, such as those involving surging and collapse of ice sheets, but models that invoke degassing of oceanic CO₂ alone as the prime cause of warming find it difficult to explain the rapidity of change, the rise in CH₄ and the ¹³C age plateau. Emission of CH₄, especially from western Siberia, may have been a critical factor in triggering deglaciation.

E. NISBET

Department of Earth Sciences,
University of Cambridge,
Cambridge CB2 3EQ, UK

1. MacDonald, G.J. *J. Petrol. Geol.* **5**, 341–362 (1983).
2. Nisbet, E.G. *Can. J. Earth Sci.* **27**, 148–157 (1990).
3. Fairbanks, R.G. *Nature* **342**, 637–642 (1989).
4. Raynaud, D. *et al. Nature* **333**, 655–657 (1988).
5. Neftel, A. *et al. Nature* **331**, 609–611 (1988).
6. Oeschger, H. *et al. Phil. Trans. R. Soc. A* **323**, 45–56 (1987).
7. Chappellaz, J. *et al. Nature* **345**, 127–131 (1990).
8. Stouffer, R.J. *et al. Nature* **342**, 660–662 (1989).

Bird migration

SIR—Recent experimental work carried out in Germany^{1,2} has shown that 'migratory restlessness' (*Zugunruhe*) in the Blackcap (*Sylvia atricapilla*) and other passerines has a strong genetic component, thus lending support to the widely held view that so-called partial migration in birds is mainly under genetic control. As discussed by J. Greenwood in a recent News and Views article³, partial migration may be maintained by natural selection as a mixed evolutionarily stable strategy (mixed ESS), with the two strategies having equal pay-offs at equilibrium. Here we argue instead that partial migration could be a conditional strategy in which external, non-genetic factors strongly influence the migratory behaviour of individuals.

Although Berthold's experiments did show that timing and direction of migration were very similar in caged and free-living populations, nothing is known about the link between migratory restlessness in the laboratory and the actual decision making by an individual under natural circumstances. Moreover, recent theoretical work (see ref. 4 and references therein) has shown that a strong genetic component in the migratory trait is no unambiguous proof for partial migration

being a mixed ESS, and that partial migration may well be a conditional strategy in which balanced pay-offs are not necessary. Migratory individuals may simply make the best of a bad job. But the evidence to distinguish between these hypotheses is lacking⁴.

The partial migration of the European robin (*Erithacus rubecula*) also has a strong genetic component in captivity⁵. Our work in north Belgium⁶ has shown that most residents are males, whereas other males and most females migrate south. We found that local survival of resident adult males was on average 2–3 times higher than that of migratory individuals (50 per cent versus 17 per cent), and that even during the extremely cold winter of 1984/85, survival of residents was not lower than that of migrants. Resident males were almost twice as likely as migratory individuals to obtain a partner (74 per cent versus 44 per cent). Combining survival and mating success shows that the expected reproductive success of migratory males is 2–4 times lower than that of resident males. It seems highly unlikely that a migrant's lower probability of surviving and mating can be compensated by an equivalently higher breeding success: migratory males arrive later in the season and cannot extend their breeding season in summer because they have to prepare for migration. Our results also demonstrate the conditional character of the migratory trait: females migrate more than males, migratory and resident males tend to use different habitats, and young of early broods are more migratory than those of later broods.

We think, therefore, that partial migration in the European robin is a conditional strategy with unequal pay-offs, where individuals are forced to migrate unless they can find a territory in which they have a good chance to survive the winter. Therefore, socially dominant individuals are more likely to be resident.

This conclusion does not contradict the strong genetic basis of migratory restlessness in the species⁵, but it does show the importance of external factors (habitat, social status and so on) in the final decision making. This conditional-strategy hypothesis could be valid for partial migration in general⁴, but, given the robin's peculiar winter social system and/or the possible multiple evolution of migratory behaviour³, this remains to be shown.

FRANK ADRIAENSEN
ANDRÉ A. DHONDT
ERIK MATTHYSEN

University of Antwerp,
U.I.A.—Department of Biology,
B-2610 Wilrijk, Belgium

1. Berthold, P., Mohr, G. & Querner, U. *J. Orn.* **131**, 33 (1990).
2. Berthold, P. *J. evol. Biol.* **1**, 195–209 (1988).
3. Greenwood, J.J.D. *Nature* **345**, 209–210 (1990).
4. Lundberg, P. *Trends evol. Ecol.* **3**, 172–175 (1988).
5. Biebach, H. *Auk* **100**, 601–606 (1983).
6. Adriaensen, F. & Dhondt, A.A. *J. anim. Ecol.* (in the press).

The law beats Maxwell's demon

SIR—John Maddox' review of the Maxwell's demon case (*Nature* 345, 109; 1990) provokes the comment that the cited approaches to the problem are misguided.

First, it is not clear what the rules of the game are for the demon. If the demon has to obey the laws of physics, then what are the arguments about? If it is immune from the physical law, then why the physical implication of the demon's computational effect (energy consumption)? And what physically interesting conclusions could possibly be drawn by considering an agent operating outside physical law?

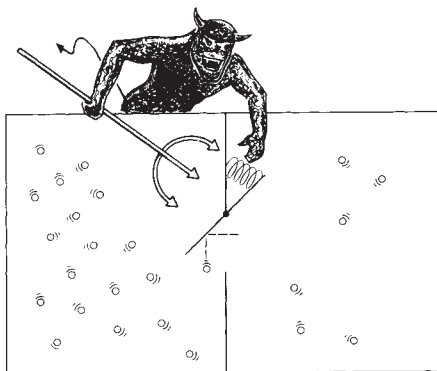
Second, if we settle for a mechanical demon — one that obeys the canon of physics (dynamics) — it will contain some moving parts; physically, the demon is a system with some physical degrees of freedom. But classical statistical mechanics postulates that, for a system in thermal equilibrium, the energy is equally partitioned among all the degrees of freedom. Because the demon is coupled to the system, it would get its share of the equipartitioned energy. The question is, then, would this fluctuating infusion of energy nullify the demon's anti-second law effort? It could be retorted that the objective of the demon is to prevent a thermodynamic equilibrium in the first place, but then to what does the temperature T in the expression $kT \log 2$, the energy of computing, refer?

As an illustration, consider a simple example of a box partitioned into two equal halves by a wall, the halves containing an equal number of molecules. Suppose there is a small, almost massless, door in the wall which opens only from one side when struck by a molecule from that side. Suppose further there is a small spring attached to the door which automatically closes the door after a molecule has passed through it. That is our mechanical demon.

With this arrangement, should we not expect more molecules be assembled in one half rather than in the other, making a pressure gradient that could be used for extracting work in contradiction to the second law of thermodynamics? It could be argued that the kinetic energy acquired by the door through equipartition keeps it in motion, allowing molecules to pass in both directions with equal probability. Furthermore, to stop the door from recoiling after being slammed, a dissipative stopping mechanism would be required, whence any gain in negative entropy through a pressure gradient would be offset by an increase of entropy through the dissipation process of the slamming door.

This raises a third point. No assessment of the energy required for processing bits of information can be based on Maxwell's demon-type arguments. An expression like $kT \log 2$ makes reference to the

statistical concept of temperature. Clearly, one cannot substantiate such a formula through examples with only one or two molecules, which is more like the situation of zero temperature, at which entropy and pressure are assumed to be zero anyway. Instead, to compute the energy needed for processing information, one has to make a



physical model of the computer (and the environment) and apply basic physics (dynamics).

My last point is about dissipation. If the mechanical system is only "almost frictionless", there is still dissipation, which in the long run builds up entropy. And if it is assumed to be completely frictionless, it will become severely agitated through equipartition. Also, the Poincaré recurrence theorem would apply; that states that the system recurs arbitrarily close to its initial state after a time long enough. Thus the 'second law of thermodynamics' would have to be conditioned by the estimate of the probability of finding the system in the state of maximal entropy.

Finally, I would put the demon's problem this way: as a cognitive being (like Laplace's demon), the demon has all the freedoms and power of heaven and hell, but as soon as the demon tries to temper with earthly matters, it is under the jurisdiction of physical law.

FRANK BORG

Palo 15,
SF-67400 Karleby,
Finland

SIR—Maddox (*Nature* 345, 109; 1990) reverts to Maxwell's demon and subsequent discussions by Szillard and others concerning its practical use to circumvent the second law of thermodynamics and elude energy constraints. Can the demon in theory defend or resolve the Boltzmann paradox of purporting to deduce a macro time-asymmetric one-sided increase in entropy from time-reversible microscopic differential equations?

The answer is no. If a demon could operate as described, the resulting microscopic differential equations are non-hamiltonian. This is not so much because of some special violation of energy

conservation, but rather because the observable particles going through apertures or being reflected from trapdoors obey new differential equations that are definite, but definitely not time-symmetric. There is no paradox to be resolved when time-asymmetric microscopic equations entail macroscopic measurements that are time-asymmetric.

To confirm the point, consider a frictionless trough containing two perfectly elastic balls and being bounded by perfectly elastic walls at each side. Before Maxwell's interfering, the balls proceed from their specified initial positions and velocities according to Newton's simplest law of non-acceleration — until a ball hits a wall or another ball. By definition of 'perfectly elastic' entities, approximatable in real-life experiments, a ball reverses its direction instantly without losing speed just when hitting a wall; two balls that collide, if of the same mass, instantaneously exchange their velocities. When averaged over all future time, each half of the trough has the same mean number of balls. (Their temperatures average out equal, so to speak.) Of course, the history of the specified system going backwards in time obeys precisely the same kind of newtonian equations. Whatever statements I can make about 'mean temperatures' or about 'entropy' going forwards in time can be matched by a valid similar deduction going backwards in time.

But take Maxwell's demon. It changes linear differential equations such as $\ddot{y}_i(t) \equiv 0$ and impulsive reversals of velocities at walls and collisions into nonlinear differential equations that are complicated only to write down. Let $(-1, 1/2, +1)$ be respectively the values of a $y_i(t)$ when at the left wall, at the middle of the trough where the demon slides in Maxwell's trapdoor, at the right wall. Before the demon, when the y_i are away from the walls and are separated, they everywhere satisfy $\ddot{y}_i = f(\dot{y}_1, \dot{y}_2; y_2, y_1)$ where $f(\cdot, \cdot) \equiv 0$ for $y_1 \neq y_2$ and $|y_1| \neq 1, |y_2| \neq 1$. Now, when y_1 is near the trapdoor and away from y_2 , the demon will want to shepherd y_1 into, let us say, the left-hand corral. Effectively, between collisions the demon replaces Newton's laws by its own new rules

$$\begin{aligned} \dot{y}_i(t+) &= D[\dot{y}_i(t-), y_i(t)] \\ \text{where} \quad D[\dot{x}, x] &= \dot{x} \quad \text{for } x \neq 1/2 \\ D[\dot{x}, 1/2] &= \dot{x} \quad \text{for } \dot{x} < 0 \\ &= -\dot{x} \quad \text{for } \dot{x} > 0 \end{aligned}$$

To convince oneself of the time-asymmetry of the demonized system, ask where the system had been a short time before observing both balls in the left-half corral with one moving just to the left of the trough's midpoint. One cannot say whether the last ball will have just been trapped, and therefore that it was recently

at a positive y reading; or that the trapping might have happened a million years ago. The demonized system's differential equations are so time-asymmetric as to make them globally not uniquely time-invertible. Forwards in time the Maxwell scheme has uniquely defined trajectories.

Maddox may be right in thinking Maxwell's tongue was in his cheek, but his text has been mostly interpreted otherwise. At least Maxwell seems to have wanted to get people to think in terms of statistical mechanics by fabricating a contrafactual whimsy.

PAUL A. SAMUELSON

*Department of Economics,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA*

SIR—Maddox, in his discussion of the possibility of breaching the second law of thermodynamics (*Nature* 345, 109; 1990) raises an issue which deserves more than cursory attention. If Maxwell's demon could act effortlessly to allow more energetic particles to pass from one cavity A to another B, while restricting flow from B to A to less energetic particles, then one could transfer heat up a temperature gradient to cause B to be hotter than A. The question is important because our search for a pollution-free energy source would soon be solved for ever if an army of Maxwell's demons could be put to such use on a commercial scale.

Because we regard photons as carriers of energy and see them as a kind of particle travelling at the speed of light, we hardly need rely on the intelligence of Maxwell's demons to open and close a shutter across a hole between the two cavities. Instead it is sufficient to place a convex mirror located in cavity A and positioned away from but facing the hole. The focusing action of this mirror will ensure that those photons transfer heat from A to B until the walls of cavity B are at a higher temperature than the walls of cavity A.

This should suffice as an apparatus which will breach the second law of thermodynamics. Textbook teaching declares that "In nature heat is never found to proceed up a temperature gradient of its own accord". From this, the textbooks advance to a statement of the law according to which it is impossible for any machine to abstract heat from the coldest body of its surroundings and convert this into useful work surplus to that needed to

power the machine. Maxwell's demon can affect that self-accord of the heat transfer and this causes one to wonder if that demon does really do any work in controlling the shutter. But with the mirror discharging this physical task, it is assuredly not doing work itself and those photons proceed up that temperature gradient by their own accord as they bounce from the mirror surface. In principle, therefore, there just has to be a failure of the second law of thermodynamics.

Maddox, in his editorial on mechanical engines driven by light (*Nature* 342, 13; 1989), suggested that "it would be more than just fun if somebody were to build one". The practical implications are enormous but they highlight the need to develop a miniature thermoelectric power converter which could be incorporated between cavities A and B in a stratified system containing a large superficial heat radiating surface with numerous cavities and numerous mirror focusing elements so that a significant net power per unit volume can be generated.

What is so fascinating about this proposal is that such an energy device would not be subject to the Carnot efficiency limit. All the heat fed into the system to sustain the temperature of cavity A would emerge as electricity even though it might cycle several times between cavity A and cavity B, going one way thanks to the focusing power of the mirror and the other way as heat 'loss' through the thermoelectric converters. The efficiency of the heat-to-electricity conversion of the thermoelectric power converter is not a factor limiting what has just been said. It is just that the greater this efficiency, the smaller the volume of the structure needed for a given power output and so the smaller the capital expense incurred.

H. ASPDEN

*Department of Electrical Engineering,
University of Southampton,
Southampton SO9 5HN, UK*

SIR—Maddox's article (*Nature* 345, 109; 1990) does not mention two issues which, it seems to me, are inseparable from any such discussion.

First, the traditional assumption is that the partition operated by the demon must be frictionless so that no energy is lost as heat during operation of the partition. Maddox does not mention that the partition must also be massless if work is not to be done in opening and shutting the partition. But how can a massless partition impart momentum to those molecules whose passage it seeks to prevent? (Conversely, if the partition has no rest mass, then it must be travelling at the speed of light relative to the demon, creating an absurd situation.)

Second, I recognize that Maddox's discussion was intended to focus on the statistical interpretation of the information

theoretical aspects of the problem. The discussion is based entirely upon classical concepts. But what of the quantum-mechanical implications of the demon's ability to 'see' which molecules are travelling faster (or slower) than the average? To 'see' a molecule, the demon must exchange quanta with the molecule, thereby altering the states of both the molecule and the demon in accordance with the uncertainty principle. The assumptions underlying Maxwell's demon are *a priori* inconsistent with quantum mechanical reality.

LEE H. KRONENBERG

*Lee Biomolecular Research
Laboratories Inc.,
11211 Sorrento Valley Road,
San Diego, California 92121, USA*

No ecosystem shift

SIR—Fanning¹ suggested that atmospheric inputs of fixed nitrogen have shifted the surface ocean waters, in parts of the North Atlantic and North Pacific, from nitrogen to phosphorus limitation. Such a change represents a large perturbation of these ocean ecosystems and requires investigation. Here we wish to make two points that we believe argue that the case for such a shift is not proven.

Several years of monthly measurements of nitrate and phosphate concentrations are available from station S in the Sargasso Sea (ref. 2 and references therein; A.H.K., unpublished data). This area lies downwind of the North American continent in an area where atmospheric inputs might be expected to have their greatest effect, although we have previously argued that on an annual average their magnitudes are small compared to internal recycling³. We have analysed recent data for the period April 1985–November 1988 using the same criteria as Fanning, because our methods and detection limits are similar to those reported in the

Class	Number of samples	% of total
Nitrate undetectable	126	24
Phosphate undetectable	46	9
Nitrate and phosphate undetectable	307	59
Nitrate and phosphate detectable	42	8

521 water samples collected at station S (32° 10'N, 64° 30'W) in the upper 100 m (approximately the average euphotic zone depth) April 1985–November 1988 have been classified as only nitrate undetectable, only phosphate undetectable or nitrate plus phosphate undetectable using the criteria presented by Fanning (1). An additional class in which nitrate plus phosphate are both detectable covers mostly the period of winter overturn.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references. □

GEOSECS and Transient Tracers in the Ocean (TTO) studies considered by Fanning. Our results are presented in the table and suggest that the surface waters of this area of the Sargasso Sea are deficient in both nitrogen and phosphorus for most of the time, rather than phosphorus-deficient as suggested by Fanning. The two data sets are different in that one provides geographical coverage and the other is a time series at a single point, but we would argue from the station S data set that the situation suggested by Fanning is not necessarily the general case.

A more fundamental point, however, involves the extrapolation of nutrient detectability, as described above and in Fanning's paper, with nutrient limitation because there is clear evidence that high rates of primary productivity can be sustained at very low nutrient concentrations by intense nutrient recycling⁴. Therefore considerations of nutrient limitation based on rates of nutrient consumption and carbon production would be more appropriate.

T. JICKELLS

School of Environmental Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK

A. H. KNAP

R. SHERRIFF-DOW

Bermuda Biological Station,
Ferry Reach, GEO1, Bermuda

J. GALLOWAY

Department of Environmental Sciences,
University of Virginia,
Charlottesville, Virginia 22903, USA

1. Fanning, K. *Nature* **339**, 460–463 (1989).
2. Bigg, G.R., Jickells, T.D., Knap, A.H. & Sherriff-Dow, R. *Oceanologica Acta* **12**, 437–442 (1989).
3. Knap, A.H., Jickells, T.D., Pszeny, A. & Galloway, J. *Nature* **320**, 158–160 (1986).
4. Goldman, J.C., McCarthy, J.J. & Peavey, D.G. *Nature* **279**, 210–215 (1979).

Genetic basis of grammar defect

SIR—Vargha-Khadem and Passingham¹ and Fletcher² in their responses to my Scientific Correspondence present no data or arguments to challenge my hypothesis³ that there is a deficit in the ability of my subjects to construct an underlying grammar for abstract morphemes like number and tense. They suggest that other subjects in this heterogeneous clinical category have other problems. That is to be expected, but it is also true that the kind of errors that I cite are widely reported in the literature. They say that the subjects in the family in question make additional kinds of errors, some of which I would dispute, but they fail to show that they are relevant to the hypothesis which I have proposed.

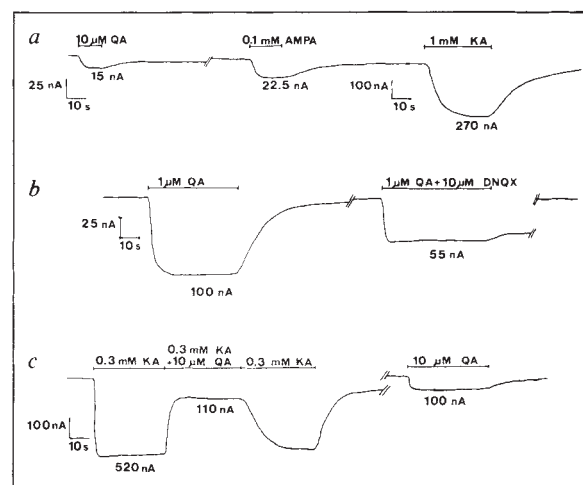
What really is at stake here are not the data, but rather how to account for them. My explanation hypothesizes that the subjects are impaired because they cannot

How many NMDA receptors?

SIR—The distinction between glutamate receptors coupled to ion channels is based on their selective activation by different agonists. Three main subtypes have thus been defined: NMDA, kainate and

response suggests the existence of two separate non-NMDA receptors.

We have further characterized the GluR-K1 clone expressed in oocytes and find that it indeed responds to kainate



Currents induced in *Xenopus* oocytes injected with GluR-K1 transcripts. Records from oocytes clamped at -60 mV. *a*, Responses to quisqualate (QA), kainate (KA) and AMPA (5 ng transcripts). *b*, Antagonism of QA by DNQX. *c*, During a prolonged application of KA, QA reduced the KA response to the level attained with QA alone. Traces *b*, *c* are from the same oocyte injected with 20 ng transcripts. Technical details available from J.R. on request.

quisqualate receptors^{1,2}, although the last two may activate the same receptor³. The recent report by Hollmann *et al.*⁴, of the cloning of GluR-K1 complementary DNA encoding a protein of relative molecular mass 100,000 which, after *in vitro* transcription and expression in *Xenopus* oocytes, induces kainate but no quisqualate

but also to quisqualate (see figure). Responses induced by kainate and quisqualate are blocked by DNQX, and GluR-K1 has a higher affinity for quisqualate than for kainate. (The maximally effective quisqualate concentration elicited only about a tenth of the maximal response to kainate.) During a prolonged application of kainate, addition of quisqualate reduces the recorded current to the level obtained with quisqualate alone (see figure). Currents induced by kainate, quisqualate and domoate display similar sensitivity to membrane potential and show pronounced inward rectification.

Our demonstration that a single protein expressed from a single complementary DNA responds to quisqualate and kainate with the same current-voltage curve supports the hypothesis that the *in vivo* responses to these agonists are mediated by the same non-NMDA glutamate ionotropic receptors.

BERTRAND LAMBOLEZ

PASCAL CURTCHET

JACQUES STINNAKRE

PIOTR BREGESTOVSKI*

JEAN ROSSIER

LIA PRADO DE CARVALHO

Centre d'Etudes du Système Nerveux,
CNRS, 91198 Gif sur Yvette,
Cedex, France.

*Permanent address: USSR Cardiology Research Centre,
Academy of Medical Sciences, Moscow USSR.

1. Mayer, M.L. & Westbrook, G.L. *Prog. Neurobiol.* **28**, 197–276 (1987).
2. Verdoorn, T.A. & Dingledine, R. *Molec. Pharmacol.* **34**, 298–307 (1988).
3. Rassendren, F.A., Lory, P., Pin, J.P., Bockaert, J. & Nargeot, J. *Neurosci. Lett.* **99**, 333–339 (1989).
4. Hollmann, M., O'Shea-Greenfield, A., Rogers, S.W. & Heinemann, S. *Nature* **342**, 643–648 (1989).

construct language rules at the necessary level of abstraction⁴. Your correspondents^{1,2} prefer an explanation which hypothesizes that the subjects are impaired because they cannot hear or cannot pronounce certain sounds, and this indirectly affects the grammar carried by these sounds. A wide range of data can be explained by my hypothesis but not by theirs. The data clearly show that it is not the marker *ed* or *s* which is impaired, but the underlying grammatical categories of number and tense. Moreover, the data confirm that all such features are impaired, even those like aspect, for which the surface marking is produced.

There is now evidence from twin studies⁵ and from large statistical studies⁶ which is consistent with the hypothesis that developmental language impairments may be linked to an autosomally dominant

gene. The exciting question is what does this gene do? Does it affect some peripheral aspect of language processing which only indirectly has consequences for language, or does it directly affect the capacity to acquire language? The abstract morpheme hypothesis supports the latter view and provides a more complete account of the data.

M. GOPNIK

Department of Linguistics,
McGill University,
Montreal, Quebec,
Canada H3A 1G5

1. Vargha-Khadem, F. & Passingham R.E. *Nature* **346**, 226 (1990).
2. Fletcher, P. *Nature* **346**, 226 (1990).
3. Gopnik, M. *Nature* **344**, 715 (1990).
4. Gopnik, M. *Language Acquisition* (in the press).
5. Borges-Osorio, M.R.L. & Salazano, F. *Acta Geneticae med. gemell.* **34**, 95–100 (1985).
6. Tomblin, J.B. *J. Speech Hear. Disord.* **54**, 287–295 (1989).

Too many mothers?

W. F. Bynum

Charles Darwin. A New Biography. By John Bowlby. *Hutchinson: 1990. Pp. 511. £19.95, \$24.95.*

"BLOOD will have blood", said Macbeth. "Biographies will have biographies", say I. There is nothing more certain to permit the prediction of a new biography of a scientist (or anyone else, for that matter), than the existence of several other ones. Of the writing of biographies of Darwin, there seems to be no end. To Ronald Clark's serviceable one of 1985 and Peter Brent's underrated one of 1981 (neither cited in the present volume), we can now add John Bowlby's full-length study of Darwin's life and health, a subject the US psychiatrist Ralph Culp tackled in similar fashion in 1977. There are at least two more biographies in the pipeline, and the magisterial *Complete Correspondence* has now reached five stout volumes*, with three times that many more promised.

Darwin, then, continues to fascinate, partly because he is such a major scientific figure, but equally because he was such an enigmatic human being. It is this latter theme which John Bowlby assaults, in what can best be described as a domestic biography. Bowlby is a distinguished child psychiatrist who has spent much of his long career studying the effects of childhood loss — especially maternal deprivation — on the subsequent development and mental health of those so deprived. Darwin's mother died when he was eight; this, Bowlby insists, is the key to the adult Darwin's life of hypochondriasis, low spirits and valetudinarianism, to his constant fretting about his children's health, his frequent consultation with London specialists and regular pilgrimages to the spas which were so important to the anxious Victorian middle classes. Further, Bowlby asserts, Darwin's relationship with his father, Dr Robert Waring Darwin, was coloured by the fact that the elder Darwin's personality was at least partly moulded by his own maternal loss. Change the parental sex, and one might change the Shakespearian illusion, to Claudius reminding Hamlet:

But, you must know, your father lost a father,
That father lost, lost his: and the survivor
bound

In filial obligation for some term

To do obsequious sorrow.

Hamlet, we are told, was oedipal; Darwin, it seems, was simply chronically depressed, many of his physical symptoms — nausea, retching and vomiting, dizziness

and lethargy — caused by hyperventilation syndrome. To be fair, Bowlby sticks close to his texts: Darwin's diaries, *Autobiography* and letters. This biography contains refreshingly little psychiatric or psychoanalytical jargon. It is a domestic study rather than a psychohistory. Furthermore, it is eminently readable and

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Darwin — A life of one long struggle?

reasonably accurate. It will not harm anyone and might even encourage those who pursue it to enquire further into Darwin's achievements. Nevertheless, Bowlby's is only a partial and selective study of the laird of Down House, and suffers from two main failings, regardless of whether one accepts the book's principal premise.

First, for a "new" biography, Bowlby bases his work too closely on printed sources. For the first half of Darwin's life, this does not matter too much, for the *Correspondence* compensates. Besides, one might say, Francis Darwin published a lot of his father's correspondence in his three-volume *Life and Letters*, to which he added a further two volumes of

*The fifth and most recent volume of Charles Darwin's correspondence was reviewed in *Nature* by John Hedley Brooke on page 523 of the 9 August 1990 issue.

pure correspondence. Darwin's daughter Henrietta edited a further couple of volumes of domestic letters from her family's incredibly rich archives. There is enough to carry us along, but the last half of Darwin's life here becomes too much of a selective rehash of Francis and Henrietta. One of the things which the modern *Correspondence* shows is simply how much more there is.

The other lacuna, acknowledged by Bowlby, is in the assessment of Darwin's scientific milieu. Given the author's warning, readers should look elsewhere for an analysis of the scientific content of Darwin's work. Nevertheless, even on its own terms Bowlby's account of Victorian natural science is dated and at times misleading. Easily available scholarship on the geological community, on the reception of *The Origin*, on scientific naturalism, and on specific aspects of Darwin's research, could have saved Bowlby from oversimplification or error. The bibliography is honest but not especially impressive.

That said, the portrait we are presented of Darwin is not unconvincing. Most recent commentators on Darwin's health have opted for psychological or psychosomatic explanations, and his childhood experiences may well have been crucial in the formation of his personality. Bowlby takes the fact that Darwin remembered so little about his mother, and never referred to her death when consoling others or experiencing subsequent losses himself (especially the painful death of his eldest daughter Annie when aged 10), as evidence that he was never able as a young boy properly to grieve his mother's passing. He was subsequently mothered by two of his elder sisters and by a favourite nanny. He married Emma Wedgwood a few weeks before his thirtieth birthday; she then took over the maternal role. Cynics might argue that Darwin had too many mothers, not too few.

Mother or mothers apart, Darwin did present an anxious and depressed persona during most of his adult life. He worried about money, fretted over Emma's regular pregnancies and dreaded the possibility that his children had inherited his own weak constitution. Some of the chapter titles from Bowlby's biography tell their own tale of woe: "a tragic loss"; "a vulnerable personality"; "a troubled spirit"; "a bitter mortification"; "grief never wholly obliterated"; "a fearful disappointment". Darwin has made it easy for us to view his life, particularly after he returned from the *Beagle* voyage, as one long struggle.

There is a paradox in this, though. Are we really to take completely at face value this tale of almost unremitting woe? Viewed from another angle, Darwin's is a

Mary Evans Picture Library

rather different story: of one born into comfort and financial security who refused to lead a life of selfish ease; of one, marrying late, who produced a bevy of healthy and distinguished children; of a man with a seemingly endless capacity for close and rewarding friendship; finally, of a man, plagued by ill-health from whatever source, who produced a dozen books, most of which went through revisions, on a breathtaking range of topics. Several of Darwin's doctors (most of whom he outlived) are remembered today primarily because he consulted them. Maybe it is time someone wrote a cheerful biography of this greatest of all naturalists. □

W.F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

■ Alvar Ellegård's *Darwin and the General Reader* has just been published by the University of Chicago Press. The author analyses the reception of Darwin's theory of evolution by the British press between 1859 and 1872. Originally published in 1958, this new paperback edition costs \$20.75, £14.25. □

Icy disputes

John H. McD. Whitaker

James David Forbes, Pioneer Scottish Glaciologist. By Frank Cunningham. *Scottish Academic Press: 1990. Pp.329. £25, \$49.95.*

THIS book describes the life of James David Forbes (1809–1868). Forbes had a difficult childhood with an over-protective father who insisted that he should study law to become an advocate, but he nevertheless worked on the polarization of heat, meteorology, magnetism and other scientific topics. He became professor of natural philosophy at Edinburgh at the age of only 24, a brilliant alpinist when his health allowed, and at 51, principal of the United College of St Andrews (where the author had access to his original notebooks, journals and voluminous correspondence). It was Forbes' enthusiasm for mountaineering in the Alps and hearing a lecture by the Swiss glaciologist Louis Agassiz which aroused his interest in the problems of glacier motion.

In August 1841, Forbes joined Agassiz on the Unteraar Glacier. Soon, he was noticing a vertical disposition of white and blue bands in the ice which could be seen to penetrate deeply into the glacier. Agassiz was dismissive of these bands, but when Forbes published an account of his 'ice structure', Agassiz claimed them as his own discovery, not admitting that he had heard a paper on this topic read, but not published, by Guyot two years earlier,

which Forbes had not. This wrangle over priority of discovery estranged Agassiz and Forbes for life. Thereafter, Forbes devoted all his field research to surveying Alpine glaciers, especially the Mer de Glace in the Mont Blanc region and — understandably — publishing his results as fast as possible to forestall Agassiz, by writing many *Letters on Glaciers* while still in the field.

Over a lamentably short active time on

plagiarized from the writings of Bishop Rendu, an accusation widely disseminated in Tyndall's book *The Glaciers of the Alps*. In 1859, Agassiz not surprisingly supported Tyndall against Forbes. Forbes and his supporters, of course, replied to these unfounded allegations of dishonesty by upholding the honour of Forbes but missed the opportunity, Cunningham maintains, of restating his very positive achievements in glaciological research.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Forbes (left) with his assistant in the natural philosophy classroom, Edinburgh University.

field work of just 10 years (he suffered and died from tuberculosis), Forbes made significant contributions to theories on the structure and movement of glaciers; for example he discovered that a glacier moves continuously, but more slowly in cold weather and at night; that the central part moves faster than the sides; and that the rate of advance is not uniform in different parts of the same glacier. Forbes also maintained, against much opposition, that "the manner of movement of the surface of a glacier coincides with the manner of motion of a viscous or semi-fluid body". Today, a century and a half later, glacier behaviour is best termed visco-plastic or as a generally nonlinear viscous incompressible fluid, so Forbes was very near the mark. Detractors at the time mistakenly put much faith in the process of regelation.

A leading opponent was Tyndall, the mountaineer who was beaten to the top of the Matterhorn by Whymper's ill-fated expedition, whom Cunningham — a master of trenchant phrases — describes as "a man with a cantankerous streak" and who "developed a particular antipathy against Scottish scientists", as well as, of course, Whymper. Tyndall, aided by T. H. Huxley, in 1856 set out to persuade the scientific public that Forbes' better ideas, to which they largely subscribed, had been

This led to the neglect of Forbes' results and lack of recognition by later authors up to the present. Even the 'Forbes Bands' that he described and measured have been inappropriately renamed 'ogives'.

In his book, Cunningham, himself a geographer and mountaineer, sets out to redress this injustice and has succeeded in reinstating Forbes as a pioneer Scottish glaciologist who, but for the ill health that terminated his active and strenuous surveying of glaciers in 1851 at the age of only 42, would undoubtedly have made many more important discoveries. Cunningham gives a literal blow-by-blow account of Forbes and his guides hammering nails and stakes into glaciers and triangulating their positions day by day in all weathers, as well as measuring ablation, 'dirt bands' and crevasse patterns.

Cunningham also emphasizes the slow acceptance of the 'Glacier Theory' advanced by Agassiz and his Swiss colleagues. British geologists, Forbes amongst them, were very reluctant to accept that striations, *roches moutonnées*, erratics, trim lines and drift deposits had been formed by land ice that had formerly extended over Scandinavia and much of Britain. They followed Lyell in attributing striations to icebergs floating in a sea of much higher level than at present, from which they thought drift was formed as a marine

deposit. Eventually, Forbes, Ramsay, Whittlesey, Croll and others accumulated so much evidence that the 'Glacial Theory' which we take so much for granted today, became accepted by most geologists.

Not the least fascinating parts of the book are Cunningham's descriptions and assessments of the complex characters of Forbes and his friends and enemies. Forbes is described as "aloof, meticulous, cautious, exceedingly well-organised, the model of an inductive scientist", but generally disliked by his students. The ebullient Agassiz was "a man of genius, . . . capable of making leaps in the ladder of his ideas, skipping rungs in a fashion that to Forbes would have been offensive, to arrive at a new and unfamiliar position from which he would then search for data to justify his whereabouts". Tyndall did "a great deal of peering through discoloured spectacles". But in extolling Forbes' achievements, Cunningham does not spare criticism where this is due.

The book is well researched and written, but the author is at such pains to develop his arguments that there is some unnecessary repetition. It is also rather

longer than necessary because of the exhaustive details of Forbes' glacier measurements and the inclusion of some of his exciting mountaineering feats where glaciology was not much in evidence and which glaciologists (at least) could do without.

I would have liked a little more discussion at the end on the latest ideas on glacier structure and movement, and would have appreciated rather more detailed maps of Alpine glaciers and the areas of Norway that Forbes visited in 1851. The Swissair photographs of the Alps are essential to the text, which is compulsory reading for glaciologists, glacial geomorphologists and historians of science (many famous men of the mid-nineteenth century like Lyell, Murchison and Darwin are drawn into the narrative). It will also be of interest to Alpine mountaineers and to anyone who likes to read about the bitter controversies of past times and their adverse effects on those involved. □

John H. McD. Whitaker was formerly Editor-in-Chief of Geology Today and is at 11 Guilford Road, Leicester LE2 2RD, UK.

Spinning stars

C. Martin Gaskell

Pulsar Astronomy. By A. G. Lyne & F. Graham-Smith. Cambridge University Press: 1990. Pp.274. £40, \$69.50.

IT WAS ON 6 August 1967 that the first pulsar (PSR 1919+21) showed up on a chart-recording at Cambridge University. In the intervening 23 years, pulsars and neutron stars have taken their place in a very wide range of astrophysics. *Pulsar Astronomy*, the first monograph on the subject for more than a decade, provides an excellent introduction to what is known about pulsars as well as a discussion of some of the other areas of astrophysics influenced by pulsars.

In their preface, the authors say that their book will not be needed by the small band of pulsar specialists but that it is intended to be useful "perhaps to students entering the field and perhaps to those who enjoy an exciting story of discovery and of the application of basic physics to a realm unattainable in earthbound laboratories". The book is probably best suited for graduate students starting research on pulsars and for other scientists who want a quick overview of what is known about pulsars. It is about the right length and level for a one-term graduate-level introduction to the subject. Many parts do indeed tell "an exciting story of discovery" which will be appreciated by more general readers, even those with little scientific background.

Pulsar Astronomy is based in part on Sir

Francis Graham-Smith's earlier monograph, *Pulsars* (Cambridge University Press, 1977). Less than half the original text and figures are retained, but for convenience I will refer to the 1977 book as the first edition and this volume as the new edition. In the 13 years between the two there has been much observational and theoretical progress, which is readily apparent if one compares the first and new editions. Almost every chapter has undergone major updating (and the order has been changed). There are now new chapters on binary and millisecond pulsars (now being discovered at a rapid rate in globular clusters) and X-ray binaries and bursters. Our knowledge of almost every aspect of pulsars has increased greatly, as has the known number of pulsars (incidentally, a useful feature of the book is the

inclusion of an up-to-date catalogue of the 450-plus pulsars known up to March 1989).

The biggest theoretical advance since the first edition has probably been in addressing the question "why do pulsars pulse?"; the chapters discussing this problem have been totally rewritten. In the mid-1970s, there was no consensus as to whether the emission came from the magnetic poles of the pulsar or from near where the magnetic field was being swept round at almost the speed of light (the outer magnetosphere gap), but researchers were trying to explain the entire pulsed spectrum based on one or other of these origins. It is now accepted instead that there are two different origins: the high-energy photons come from the outer magnetosphere gap, whereas the radio beam originates directly over the magnetic pole. But even after 23 years of research, as Lyne and Graham-Smith point out, "the rotating lighthouse beam itself is the least understood of all the observable properties of pulsars".

The main deficiency is in the discussion of supernovae explosions (the main and perhaps only producers of pulsars), which contains quite a few inaccuracies and significant omissions. Rather surprisingly for a book published in 1990, no mention is made of the fact that we now know that there are two types of supernovae lacking hydrogen, so-called type Ia and Ib supernovae. The reader could be left with the impression that low-mass stars barely more massive than the Sun might somehow become supernovae — the standard exploding binary white-dwarf model for type Ia supernovae is never mentioned. Neither is the fact that some type I supernovae are believed to come from young massive stars (and might even form pulsars). But these deficiencies are somewhat on the periphery of the main subject material of this book. □

C. Martin Gaskell is in the Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109-1090, USA.



From macro to micro — the polished gems on the left and the kettle 'fur' ($\times 350$) on the right are two examples of crystals taken from the book *Crystals* by Ian F. Mercer. The book contains many more delightful pictures and reveals the secrets of crystals of various shapes, sizes and origins. Published by Natural History Museum Publications, London, price is £4.95 (pbk).

Books to cut your teeth on

Gordon Peters

Oncogenes and the Molecular Origins of Cancer. By Robert A. Weinberg. *Cold Spring Harbor Laboratory*: 1989. Pp. 367. Hbk \$97, pbk \$55.

Oncogenes. By Geoffrey M. Cooper. *Jones and Bartlett*: 1990. Pp. 323. \$50. Distributed in the UK by Chapman and Hall. £34.

THE fundamental importance of proto-oncogenes in the control of cell growth and differentiation, as well as the aberrations that lead to cancer, make the study of these genes one of the most fruitful and stimulating areas of current biomedical research. Books that attract newcomers to the field, spark off new ideas or provide accessible sources of information are therefore important and welcome. Curiously, although both these volumes aspire to serve the same communities, and cover essentially the same material, they are as different in style as the major protagonists.

Bob Weinberg, in his typically eloquent preface to *Oncogenes and the Molecular Origins of Cancer*, for which he serves as editor and contributor, lyrically ponders the dilemmas faced by anyone embarking on such an enterprise. For whom should the book be written, how should it be structured, and when is the ideal time to tackle such a broad and rapidly expanding topic? A year in which the Nobel committee has acknowledged the significance of cellular oncogenes seems a fortunate choice, particularly as the recipients of the prize feature among Weinberg's coauthors. Although library shelves are laden with competent review articles and chapters in books on every aspect of the subject, many written by contributors to the present volume, an authoritative compendium has been long overdue. Cold Spring Harbor Laboratory Press has a history of producing timely texts that become popular sources of reference, and this one will be no exception.

Weinberg has done well to coordinate such an impressive list of acknowledged experts, but in reality the book is nothing more than a series of good, substantial reviews. Any multi-author text suffers from a lack of continuity, as each writer constructs a chapter that can be read in isolation, and each will have personal views on what constitutes "a statement of the ideas that shape current research in this rapidly moving field". There is inevitable repetition, the diagrams and tables are sporadic and of variable quality; despite the editor's desire to avoid "exhaustive, encyclopedic recounting", the

text is heavily referenced. Few citations are as recent as 1989, but this is to be expected in an opus of this magnitude. Some semblance of order does prevail, as the first seven chapters take us from growth factors on the outside of the cell to transcription factors in the nucleus, after a stiff and all-embracing overview by Harold Varmus. But the later chapters on DNA viruses, tumour-suppressor genes, and the closing foray into human cancer by Mike Bishop, appear as important additions rather than part of a logical scheme. Thus, while the chapters are of consistently high quality, the book as a whole is heavy going for those with only a passing curiosity in oncogenes. Weinberg's noble ambition that it should serve as a "recruiting pamphlet, a device to interest and excite those outside the field" is unlikely to be fulfilled, but any conscientious conscript should take the time to read this informative volume.

In a curious twist of fate, the monograph prepared by Weinberg's erstwhile scientific rival, Geoff Cooper, is a much more approachable work. Although it takes a more superficial and personalized view of the subject, written in a uniform style that the reader may like or dislike, it avoids the need to split the subject according to the expertise of different contributors. The benefits are a greater sense of logic and continuity, leading the uninitiated through sections dealing with the nature of the cancer cell, tumour viruses, viral and cellular oncogenes, and the normal functions of proto-oncogenes.

Although Cooper's prose is rather dry, the text is enlivened by personal credits and dates that confirm his acquaintance with the main players in the field as well as the key experiments and seminal discoveries. It is also liberally illustrated, with diagrams and tables on almost every other page. Not all these figures are necessary or informative, but most are helpful and the effect is to speed the reader's progress through the historical unfolding of the subject. References are mercifully excluded from the text, again making it easier to read, but each chapter concludes with a brief summary and a compilation of significant papers, divided according to subjects or concepts. The cognoscenti might choose to quibble with the exact choice or number of references, but there are enough to set the interested reader off on the right trail. All in all, Cooper has produced an admirable book that is above all a good read.

My advice to would-be novices is clear: cut your teeth on *Oncogenes* by Geoff Cooper before tackling the more solid fare offered by Weinberg and his illustrious collaborators. □

Gordon Peters is at the Imperial Cancer Research Fund Laboratories, St Bartholomew's Hospital, Dominion House, 59 Bartholomew Close, London EC1A 7BE, UK.

New in paperback

■ *The Family Genetic Sourcebook*, by Benjamin A. Pierce, "helps you to anticipate and understand genetic diseases and disorders". It provides an A-Z on traits and disorders as well as an explanation of the basic concepts and history of genetics. Publisher is Wiley, price \$14.95.

■ Also on the subject, *DNA Fingerprinting* by Lorne T. Kirby is a practical guide to basic principles and laboratory methods in forensic analysis, paternity testing, medical diagnostics and wildlife poaching. Publisher is Stockton, price £24.95.

■ *Fire in the Rain* by Peter Gould assesses the democratic consequences of Chernobyl. Gould traces the fallout and its ecological consequences as well as the movement of information and misinformation. Publisher is Polity, price is £7.95.

■ In *Methods of Surface Analysis*, a collection of authors provides an introduction to the main techniques and how they can be used to solve problems on surfaces, thin films, interfaces and coatings. Edited by J. M. Walls, the book is published by Cambridge University Press at £15, \$24.95.

■ The second edition of *Physics in the Life Sciences* by George Duncan introduces the principles of physics to biology and medical students. First published in 1975 by Blackwell Scientific, price is £14.95.

■ David Cohen's latest book, *Essential Psychology*, is published by Bloomsbury. Written in dictionary format, Cohen provides a history of the discipline and assesses the main controversies. Price is £5.99.

ADVERTISEMENT

AIDS TARGETED INFORMATION/ATIN

TARGETS YOUR LITERATURE SEARCH

Expand your base of knowledge about AIDS/HIV with *AIDS Targeted Information/ATIN*. Every month, *ATIN* not only provides abstracts but also in-depth evaluations of the current published scientific literature on AIDS. Written by clinicians and researchers for clinicians and researchers, *ATIN* provides an authoritative command of the world's literature on AIDS. Published by Williams & Wilkins. Indexed. 12 issues per year. \$125.

To order call
1-800-638-6423

Sponsored by the American Foundation for AIDS Research (AmFAR).

Reader Service No. 430

Ozone loss in the Arctic polar vortex inferred from high-altitude aircraft measurements

M. H. Proffitt*, J. J. Margitan†, K. K. Kelly‡, M. Loewenstein§, J. R. Podolske§ & K. R. Chan§

* NOAA Aeronomy Lab, 325 Broadway, Boulder, Colorado 80303 and CIRES, University of Colorado, Boulder, Colorado 80309, USA

† Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

‡ NOAA Aeronomy Lab, 325 Broadway, Boulder, Colorado 80303, USA

§ NASA Ames Research Center, Moffett Field, California 94035, USA

The Arctic polar vortex in winter is known to be chemically primed for ozone depletion, yet it does not exhibit the large seasonal ozone decrease that characterizes its southern counterpart. This difference may be due in part to a net flux of ozone-rich air through the Arctic vortex, which can mask ozone loss. But by using a chemically conserved tracer as a reference, significant ozone loss can be identified. This loss is found to be correlated with high levels of chlorine monoxide, suggesting that much of the decrease in ozone is caused by anthropogenic emissions of chlorofluorocarbons.

IN 1985 Farman *et al.*¹ reported that spring values of total column ozone over Halley Bay, Antarctica, had decreased considerably during the previous decade, and they suggested that certain chlorine-containing compounds could be causing the observed ozone depletion. Further studies have supported these findings²⁻⁵ and refined understanding of high-latitude ozone loss.

During the 1987 Airborne Antarctic Ozone Experiment (AAOE)⁵ *in situ* measurements of many chemical species and meteorological parameters were made from a NASA high-altitude aircraft (ER-2) on twelve flights out of Punta Arenas, Chile (53° S), at cruise altitudes of ~17–19 km. *In situ* measurements of ozone (O₃)⁶, nitrous oxide (N₂O—a chemically conserved tracer)⁷, chlorine monoxide (ClO)⁸, temperature, pressure and wind speed⁹ were made.

Potential temperature (Θ) is a measure of entropy and is calculated from measured pressure and temperature. In the absence of diabatic processes, parcels tend to move on constant Θ surfaces and mix with other parcels on that surface. Because Θ increases with altitude, and air movement is stratified somewhat by Θ , we use it here in place of altitude as a vertical coordinate. Using Θ , diabatic processes (net heating or cooling) can be more easily identified. The aircraft flights were usually to 72° S on surfaces of constant Θ (isentropic). AAOE lasted from 17 August to 22 September, covering late winter and early spring in the Southern Hemisphere. It was found that the O₃ mixing ratio at 72° S decreased by 61% at $\Theta = 425$ K (17 km) over a 30-day period⁶ and was precisely collocated with very high ClO mixing ratios¹⁰. Calculations indicated that the observed concentration of ClO was high enough to link the O₃ decrease to catalytic destruction by chlorine¹¹⁻¹³. These calculations, however, included two important assumptions: that diabatic cooling was negligible and that the vortex air was not being replenished by O₃-rich air from outside the polar jet. This view of the polar jet as a 'containment vessel' is supported in the analyses of some authors^{1,14-16} but not by others¹⁷⁻²².

Possible Arctic ozone loss has also been reported recently. Balloon data taken during 1989 suggest that ozone depletion occurred late in January at an altitude of 22–26 km over Kiruna,

Sweden (68° N)²³, and that there was a depleted ozone layer from 18 to 24 km from 24 January to 22 February over Alert, Canada (82° N)²⁴. Another indication of possible ozone loss in the Arctic winter was found by the Ozone Trends Panel²⁵, where a comparison of total column ozone data from 1979–80 with 1986–87 shows a decrease of 4–10% between 65° and 80° N during the period from November to February.

An aircraft campaign, similar to AAOE, the Airborne Arctic Stratospheric Expedition (AASE) was based at Stavanger, Norway (59° N) from 3 January to 10 February 1989, to study ozone destruction mechanisms during the Arctic mid-winter²⁶. The instrumentation used on the 14 ER-2 flights was virtually identical to that used in the Antarctic, but the aircraft flew nearer the pole in the Arctic, usually to 79° N. Although the Antarctic vortex persists well into spring, the less stable Arctic vortex usually dissipates in late winter and so the Arctic measurement campaign had to be conducted earlier in the season. Large photochemical ozone loss was not expected during this period of minimum solar radiation and, as anticipated, no large isentropic temporal decrease in ozone was found in the Arctic.

The analysis we present is based on a moving coordinate system defined vertically by Θ and horizontally by latitude referenced to the vortex boundary, which was defined as the latitude of peak wind speed as measured from the aircraft. The average latitude of the boundary is 68° N with a standard deviation of 3°. For the analysis that follows, the data will be considered in three regions: the vortex exterior (from aircraft take-off at 59° N to the boundary), the outer vortex (the first 8° of latitude inside the vortex boundary, comprising about 60% of the vortex volume) and the inner vortex (the remainder of the vortex, usually including the pole). The outer vortex receives more solar radiation than the inner vortex, and so photochemical ozone loss is more likely to occur there.

An analysis similar to that used to show that there is ozone loss outside the Antarctic ozone hole²⁷ will be used to show that there is also significant ozone loss in the Arctic vortex. Fundamental to the arguments presented is that standard gas-phase photochemistry identifies the primary source region for O₃ and the loss region for N₂O as the tropical middle stratosphere and also predicts long photochemical lifetimes for N₂O and O₃ (more than a year) for high-latitude winter²⁸⁻³¹. From this, we deduce that air parcels at high latitudes with the same N₂O mixing ratios should have very similar O₃ mixing ratios. Deviations from this behaviour thus indicate chemical ozone loss occurring by processes not included in standard gas-phase photochemistry models³².

In the following section, O₃ and N₂O data are presented and their relative behaviour during the mission characterized. In the section that follows, we argue that the vortex behaves as a 'flowing processor' and these data are evaluated in this context, and an apparent O₃ loss is identified within the vortex. The origin of the loss is shown to be chemical rather than dynamical, to be enhanced in the outer vortex, and to have occurred both during and before the mission. The O₃ and N₂O data taken outside the vortex where O₃ loss is expected to be small are compared with values found within the vortex where an O₃ loss

is identified. The loss in the outer vortex is linked with high concentrations of ClO.

O₃ and N₂O at high latitudes

There are insufficient simultaneous measurements of lower stratospheric O₃ and N₂O available to determine an accurate climatology for these trace species, and so our analysis relies solely on the measurements made during AASE. An example of mid-latitude N₂O and O₃ data is given in Fig. 1a. These data are from an AASE transit flight on 21 February 1989 at 37–39° N latitude. A nearly linear relationship between O₃ and N₂O can be seen. Although the data only include N₂O values down to 165 parts per 10⁹ by volume (p.p.b.v.), the fit to the data agrees well with the linear fit to observations taken by the ATMOS (Atmospheric Trace Molecule Spectroscopy) instrument³³ during May 1985 at 28° N and 48° S from Spacelab 3. The linear

fit to ATMOS data shown in Fig. 1a includes all data with N₂O > 50 p.p.b.v., and extends the range of the linear relation, seen in the AASE data. The potential temperature, Θ , of each data point is indicated by colour and symbol type. As is typical of the lower stratosphere (altitude < 25 km), O₃ increases with altitude (increasing Θ) whereas N₂O decreases. At mid-latitudes, O₃ peaks typically above 30 km and then decreases, but N₂O continues to decrease with increasing altitude. Maximum O₃ mixing ratios at high latitudes are between 25 and 30 km, about 5–10 km above ER-2 altitudes^{23,24}.

Figure 1b represents all data with $\Theta > 400$ K taken in the vortex exterior. Also shown are linear fits to each of the Θ bins. Except at the lowest Θ , these data are very similar to those found in Fig. 1a, but with greater variability, and with lower values of O₃ relative to N₂O. The plot in Fig. 1c shows all the AASE data with $\Theta > 400$ K that are in the outer vortex. A linear

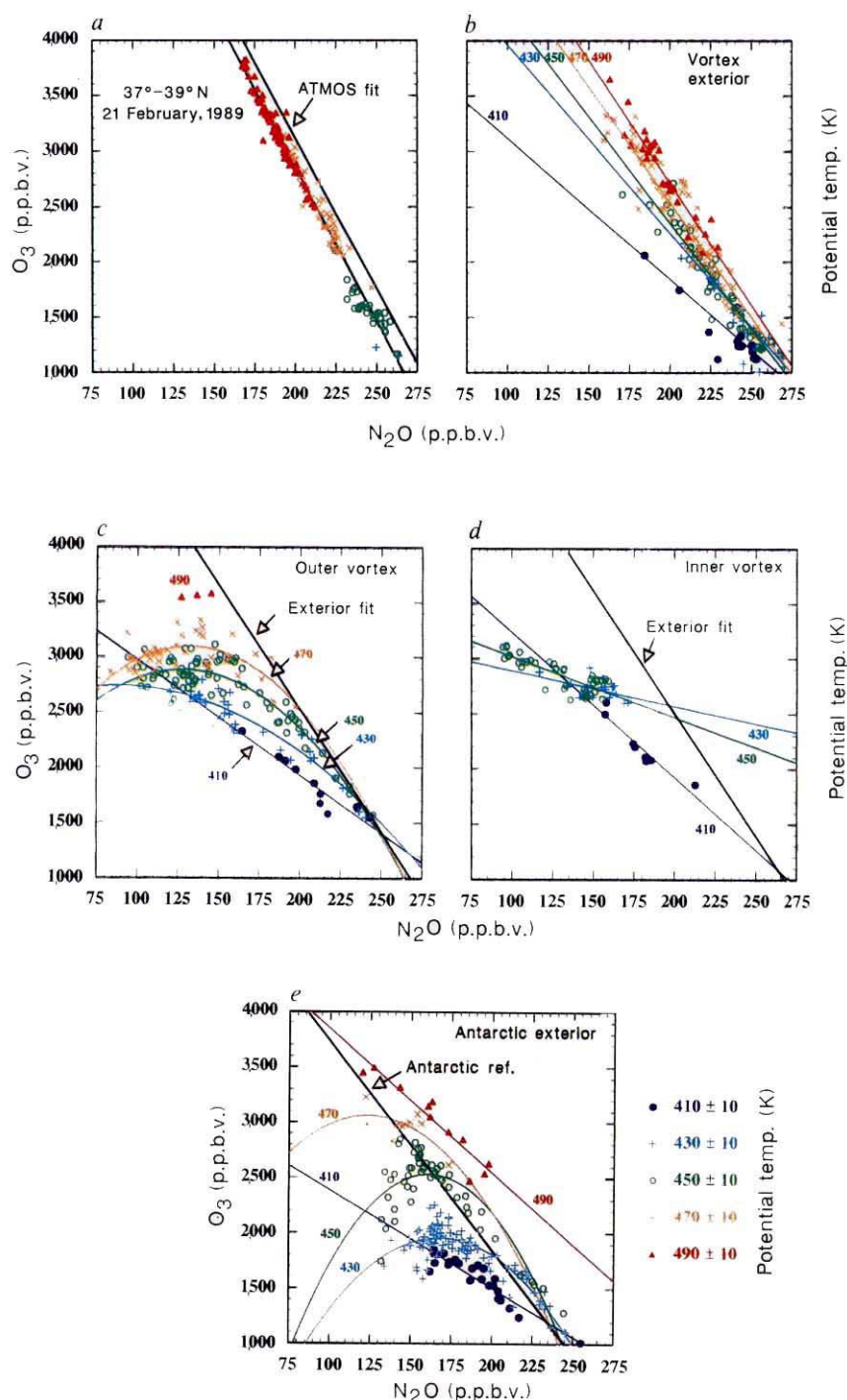
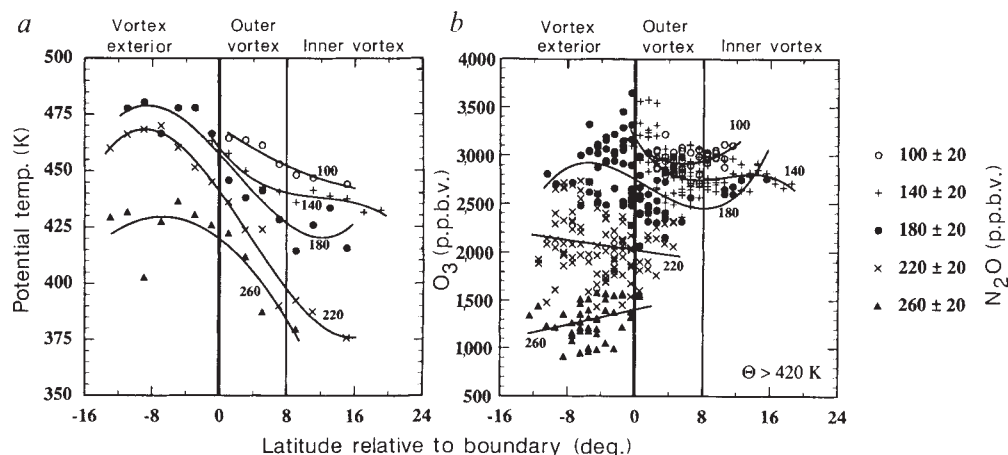


FIG. 1 O₃ is plotted against N₂O and colour coded for Θ . a, Data from 21 February 1989 at 37–39° N latitude averaged for 100 s (about 10' of latitude) and a linear fit to that data. Also shown is linear fit to ATMOS data taken from Spacelab 3 at 28° S and 40° N latitude in May 1985. b, All AASE data taken in vortex exterior (59° N to vortex boundary) with $\Theta > 400$ K, and averaged for each flight leg over 1° of latitude relative to the boundary of the vortex. Fits to data are provided for Θ bins and are discussed in the text. c, Same as (b) except that the data are from the outer vortex (the first 8° of latitude inside vortex boundary). d, Same as (b) except that the data are from the inner vortex (all of vortex interior more than 8° of latitude inside the vortex boundary). e, Same as (b) except the data are from the exterior of the Antarctic vortex during AAOE (1987).

FIG. 2 *a*, N_2O isopleths on a coordinate system of Θ plotted against latitude relative to the boundary of the vortex. Data from the entire mission are averaged over 1° of latitude relative to the vortex boundary and plotted along with polynomial fits for each of the N_2O bins. *b*, All O_3 data with $\Theta > 420$ K are plotted against latitude relative to the boundary and binned by N_2O . Data are averages over 100-s intervals (about $10'$ of latitude). A fit is provided for each bin.



fit is shown for the 410 K bin with quadratic fits for the three middle bins. The highest bin (490 K) has too few data to justify a fit. Figure 1*d* shows data from the inner vortex. Comparing Fig. 1*b–d*, the fits for the 410 K bin are all very similar, but there is little similarity in the fits to the other Θ bins. The uniformity below 420 K indicates that air is free to move from inside to outside the vortex at this level, suggesting that the 'bottom of the vortex' is where $\Theta = 410 \pm 10$ K.

Another important criterion for comparing Fig. 1*b–d*, and indicative of chemical loss of O_3 , is the deviation of O_3 mixing ratios relative to N_2O from the mid-latitude values shown in Fig. 1*a*. It is clear that for a constant value of N_2O , O_3 progressively decreases from the mid-latitudes, to the vortex exterior, to the outer vortex. In Fig. 1*b, c*, O_3 also decreases with decreasing Θ . The characteristic of decreasing O_3 with decreasing Θ for a constant value of N_2O will be referred to as ' O_3 – Θ dependence'. There is a small O_3 – Θ dependence in the vortex exterior and a larger one in the outer vortex, but none is evident in the mid-latitude data, nor in the inner vortex. The O_3 – Θ dependence in Fig. 1*b, c* is not a result of downward mixing of air from above the O_3 maximum because O_3 and Θ are negatively correlated above the maximum, whereas the O_3 – Θ dependence in Fig. 1*b, c* is positive. This implies that the O_3 – Θ dependence is not a result of dynamics, but of chemistry.

This O_3 – Θ dependence was also clearly seen in the Antarctic exterior (Fig. 1*e*) and seems much like the Arctic outer vortex (Fig. 1*c*). Also given in Fig. 1*e* is the Antarctic reference that was used for assessing O_3 loss²⁷. That reference was based on a more restricted data set than is available for the Arctic, and may underestimate O_3 loss because it was chosen very conservatively.

Isopleths for N_2O (lines of constant N_2O) within our Θ –latitude coordinate system are shown in Fig. 2*a*. Each point is

an average over 1° of latitude and over all flights of the mission. Polynomial fits are provided for each of the N_2O bins and are good approximations to the N_2O isopleths. The poleward decrease in Θ along the isopleths indicates diabatic cooling within the vortex. Diabatic cooling rates during the mission have been approximated theoretically³⁴ and from AASE N_2O data³⁵. Although these analyses do not necessarily represent average vortex conditions, they do imply that cooling of at least 0.9 K per day in Θ (0.4 K per day in temperature) occurred during the mission both inside and outside the vortex.

Figure 2*b* confirms the O_3 loss seen in Fig. 1*a–d* but in a spatial context, by plotting O_3 against latitude relative to the vortex boundary for the same five N_2O bins of Fig. 2*a*. Only data above 420 K are plotted, to exclude data that may be affected by the relatively unrestricted horizontal exchange of air at the bottom of the vortex. Again polynomial fits are included for each of the N_2O bins and terminated where the data end. In the outer vortex, O_3 decreases towards the pole, and where there are data within the inner vortex, the fits there either increase or flatten out indicating less ozone loss. The outer vortex receives more solar radiation, and so the observed spatially enhanced decrease is consistent with O_3 being photochemically destroyed.

O_3 loss in a flowing processor

Early research indicated that tropospheric air enters the lower stratosphere in the tropics and exits at higher latitudes^{36,37} and this is still generally accepted³⁰. That is, the dominant atmospheric circulation is upward at low latitudes where heating lifts air into the region of the stratosphere where ozone is produced, then poleward and at high latitudes downward, due to diabatic cooling. As previously mentioned, some authors believe that the Antarctic jet effectively terminates the poleward component, but only the question of Arctic containment will be discussed here.

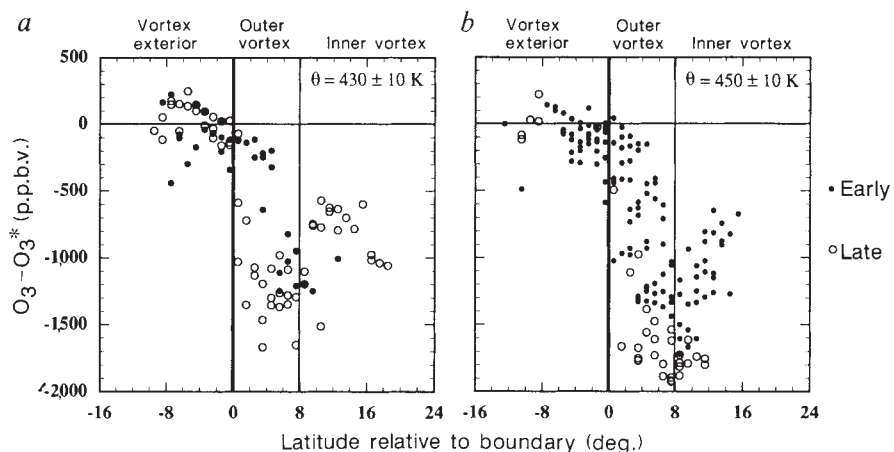


FIG. 3 $O_3 - O_3^*$ plotted against latitude relative to the vortex boundary for the early data (before 30 January) and the late data (the remainder of the mission), each point representing, from a flight leg, the average over 1° of latitude relative to the boundary. Negative values indicate loss of O_3 relative to the vortex exterior. *a*, Includes only those points with $\Theta = 430 \pm 10$ K. *b*, Includes only those points with $\Theta = 450 \pm 10$ K.

It has been shown that total column O_3 substantially increases at polar latitudes in the Northern Hemisphere throughout the winter months^{38–40}. Because O_3 is not created at high latitudes during the winter, any increase in column O_3 must be a result of poleward flux. A recent comprehensive analysis⁴¹ shows that the meridionally averaged O_3 column at polar latitudes increases by 40 to 70 Dobson Units (DU) per month from December to February, corresponding to column changes of about 12–20% per month. With regard to 1989, satellite total column data show an increase of about 15% in the outer vortex during the mission, and an increase of more than 15% is evident from 11 column measurements and 19 balloon soundings taken from 10 January to 8 February at 82° N (ref. 42). The column increase can only be explained by a poleward flux of mid-latitude air that carries more O_3 into the polar vortex than is removed by dynamical or chemical processes. We deduce therefore, both from the Arctic long-term record and analysis of 1989 data, that the vortex is not isolated, although the Arctic polar jet may restrict poleward flux.

A more difficult question to answer that is somewhat controversial is 'At what altitudes does significant mass enter the Arctic vortex?'. Leovy *et al.*⁴³ argue that diabatic descent within the vortex, along with major and minor stratospheric warming events in the middle stratosphere, account for the polar column increase, and Kent *et al.*⁴⁴ use aerosol extinction measurements to indicate that mass flow into the vortex is somewhat restricted at its boundary. These analyses may seem to imply, but they do not conclusively show, that the polar jet completely isolates the vortex in the lower stratosphere. For consider the following: (1) The analysis of Leovy *et al.* indicates that only one-third of the total column O_3 increase occurred above 30 mbar (~ 23 km). (2) Kent *et al.* infer from vertical mass-flux rates near the pole that the flux into the vortex between 14 and 24 km is approximately

equal to the downward flux within the vortex from above 24 km. (3) Only 20–25% of the Arctic winter O_3 column is above 25 km and 12–20% column increases occur each winter month. This implies that a mass of O_3 equivalent to its column above 25 km must be added to the column within the vortex every 1–2 months, plus an additional amount to balance any flow out of the bottom of the vortex. Furthermore, if (2) is accurate, then the column increase below 23 km from (1) could result primarily from O_3 flux into the vortex in the lower stratosphere rather than from vertical descent. This flux could occur during warming events⁴³ (isentropic) or from diabatic circulation into the vortex. The qualitative arguments that follow do not generally rely on either of these positions, but the arguments involved in quantifying O_3 loss do.

Searching for a possible explanation for the O_3 – Θ dependence that is consistent with Fig. 1a–d and does not require a large winter O_3 loss, we consider that the loss may have occurred before winter. Theoretical and experimental evidence for a net O_3 loss over Antarctica during the continuous daylight of the summer has been presented by Farman *et al.*⁴⁵. The authors report that the loss is not primarily due to chlorine but to low concentrations of N_2O_5 and an associated enhancement of the NO_x catalytic cycle in the middle stratosphere. These loss mechanisms apply as well to the Arctic. For the summertime loss to affect our wintertime analysis requires that a substantial part of the depleted lower-stratospheric air remains at high latitudes until the polar vortex has formed in November, and is then contained at polar latitudes until the end of the mission in February. Diabatic cooling during the winter without concurrent O_3 loss implies an isentropic O_3 increase (because O_3 increases with Θ), but this was not observed. Therefore the O_3 – Θ dependence is not simply a residual of summertime loss, but must reflect wintertime loss. Furthermore, the O_3 climatology

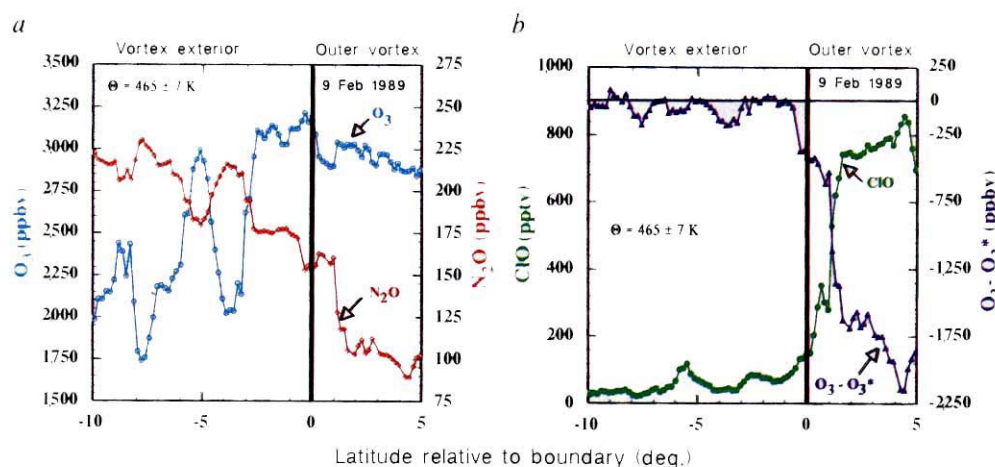
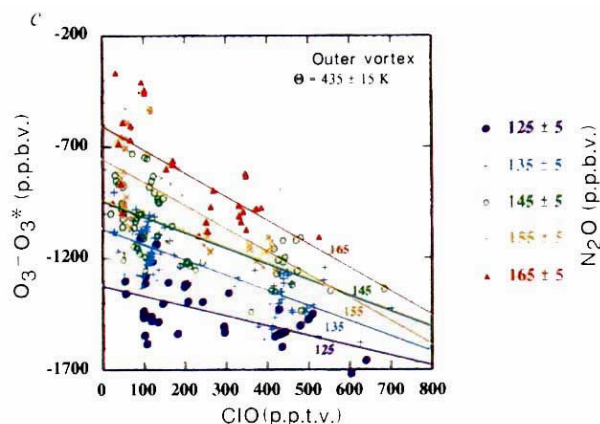


FIG. 4 a, O_3 and N_2O data from 9 February 1989 are plotted against latitude relative to the boundary along a flight leg with $\Theta = 465 \pm 7$ K. b, Same flight as in (a) except $O_3 - O_3^*$ and ClO are plotted against latitude relative to the boundary. Negative values for $O_3 - O_3^*$ indicate O_3 loss relative to the vortex exterior. c, $O_3 - O_3^*$ is plotted against ClO , including all mission data in the outer vortex with $\Theta = 435 \pm 15$ K. Data are 100-s averages that are binned by N_2O and a linear fit is provided for each bin.



discussed earlier⁴¹ shows a column increase of 12–20% per month from December to February, but also shows that the column decrease reaches a maximum in May at 12% per month, decreasing almost linearly to no change in October. Assuming the column loss in summer is representative of chemical O₃ loss, we conclude that a summer loss of O₃, terminating months before AASE, is not likely to contribute substantially to the loss reported here. But to quantify its possible effect requires very accurate O₃ and N₂O data measured simultaneously throughout the polar lower stratosphere in autumn and winter, and no such data are available.

Returning to the O₃– Θ dependence, a plausible explanation for the low O₃ and its dependence on Θ can be constructed in the context of a ‘flowing processor’. First, recall that air is diabatically cooling both inside and outside the vortex, that N₂O is chemically inert in the high latitude winter, and that mixing ratios do not change due to expansion and contraction. Therefore, if mixing of air parcels of differing N₂O content is ignored, the movement of a parcel that is diabatically cooling must be restricted to a surface described by the N₂O isopleth on which that parcel resides. This can be visualized with or without containment of the lower vortex. Consider an air parcel initially a few degrees of latitude outside the vortex that is diabatically cooling. The parcel remains on a surface of constant N₂O, so as Θ decreases, the parcel moves poleward and downward into the vortex (see Fig. 2a) thus describing a ‘flow’ into the vortex in the lower stratosphere. Alternatively, if the lower vortex is assumed to be contained by the polar jet, the cooling would be characterized as a change in N₂O on a surface of constant potential temperature, Θ (ref. 35). In this case the flow is vertical and downward as the isopleths sink. But in both cases, there is flow through the surfaces of constant Θ within the vortex in the lower stratosphere.

Let us hypothesize that by some chemical ‘process’, as an air parcel diabatically cools, O₃ is lost, and see if the resulting ‘flowing processor’ would produce the observed data. Figure 2b shows O₃ decreasing poleward along the N₂O isopleths, in the outer vortex, but not in the inner vortex. That is, the O₃ loss seems to be less in the inner vortex. Therefore our cooling/O₃-loss hypothesis, together with an enhanced loss in the outer vortex (where there is greater solar radiation), is consistent with Fig. 2, with or without mass flux into the lower vortex.

Similar reasoning can also be applied to Fig. 1b–d to again test our expanded hypothesis, this time assuming that there is mass flux into the lower vortex. Consider an air parcel with, for example, N₂O = 175 p.p.b.v. originally outside the boundary at $\Theta = 470$ K and O₃ = 3,050 p.p.b.v. (Fig. 1b). Later that parcel contains only 2,400 p.p.b.v. of O₃ when $\Theta = 430$ K inside the boundary (Fig. 1c), representing a 20% O₃ loss. But the inner vortex O₃ at 430 K is 2,700 p.p.b.v. and, as previously noted, shows no O₃– Θ dependence (Fig. 1d). Again, these data indicate that increased exposure to solar radiation in the outer vortex enhances the O₃ loss. The lack of O₃– Θ dependence in the inner vortex could be a result of enhanced isentropic mixing in that region and the low O₃ may be due to transport from the outer vortex where the loss is primarily occurring. The summertime O₃ loss discussed earlier could also explain part of the inner vortex loss.

As discussed above, air parcels at high latitudes with the same N₂O mixing ratios would be expected to have very similar wintertime O₃ mixing ratios. This was deduced without requiring the somewhat controversial assumption of significant mass flux into the vortex in the lower stratosphere. How accurately we can quantify O₃ loss from N₂O depends on how well the O₃ and N₂O reference relationship is known, and this is related to the altitude of mass flux. That is, a significant flux across the boundary at or near flight altitudes implies that air sampled inside the boundary recently entered the vortex, carrying with it the measured characteristics of the exterior. A path higher in the stratosphere requires the assumption that the O₃–N₂O refer-

ence relationship is preserved at altitudes above the region sampled by the aircraft, and therefore implies greater uncertainty in the amount of O₃ loss. Figure 1a supports this assumption.

To quantify O₃ loss within the vortex, a linear least-squares fit to all the exterior data (Fig. 1b) is calculated, excluding those points at the bottom of the vortex where $\Theta < 420$ K. This fit is shown in Fig. 1c, d and is given by

$$O_3^*(\text{p.p.b.v.}) = 7,019 - 22.42 \times N_2O (\text{p.p.b.v.})$$

where O₃^{*} approximates the O₃ mixing ratio in the vortex exterior from the mixing ratio of N₂O. For any sampled parcel inside the vortex, O₃^{*} can be calculated from its N₂O content with a negative value for O₃ – O₃^{*} representing loss relative to the vortex exterior. The mid-latitude O₃–N₂O fit (Fig. 1a) is not used as a reference, instead we use the data in the vortex exterior for the comparison. For the reasons discussed in this section, O₃ – O₃^{*} may overestimate O₃ loss within the vortex.

Figure 3 shows O₃ – O₃^{*} plotted against latitude relative to the vortex boundary for $\Theta = 430 \pm 10$ K and 450 ± 10 K. Loss in excess of 1,500 p.p.b.v. is found in the outer vortex with about two-thirds of that loss in the inner vortex. The data are also binned into early (before 30 January) and late data (the remainder). There are small temporal changes seen in the outer vortex, although Fig. 3a includes virtually no data from the first half of January because early flight legs were above $\Theta = 440$ K. Temporal changes are better represented in Fig. 3b, showing a decrease of ~500 p.p.b.v. This isentropic decrease in O₃ is a result of the temporal changes in N₂O, not absolute O₃ changes. Therefore, isentropic temporal analysis of O₃ alone will not reveal the true temporal decrease.

An O₃ decrease of 13 ± 8 p.p.b.v. per day has been reported by others³⁵ also using AASE data. The analysis was restricted to a few vertical profiles within the vortex and required a diabatic cooling rate estimated from N₂O and an assumption of strictly vertical descent. If the same cooling rates (0.9 K per day in Θ) are applied to the fits in Fig. 1c and no account is made for O₃ flux into the vortex, loss rates of from 9 p.p.b.v. per day where N₂O = 210 p.p.b.v. to 24 p.p.b.v. per day where N₂O = 130 p.p.b.v. are obtained.

As O₃ is low within the vortex early in the mission, a result of O₃ flux from above or from the exterior of the vortex without concurrent O₃ loss would be increased O₃ relative to N₂O; that is, the interior O₃–N₂O relationship would appear increasingly similar to the exterior relationship. This was not observed, again implying, as found in the discussion on summertime loss, that O₃ was chemically destroyed during the mission. From Fig. 3, O₃ loss had occurred by January. Further analysis of individual flights reveals that, on the first flight, 3 January, ~1,000 p.p.b.v. of O₃ had been lost at 465 K; that is, substantial O₃ loss had occurred before the mission.

Our analysis of the flowing processor still leaves an important question unanswered: ‘What is the cause of the O₃ loss?’. As previously mentioned, the cause of the Antarctic O₃ loss has been attributed to catalytic destruction by chlorine, and, to a lesser degree, bromine. Model calculations for the Arctic based on measured ClO and BrO suggest that O₃ loss approached 12% (about 12 p.p.b.v. per day) during the Arctic campaign⁴⁶. Coupled photochemical–microphysical lagrangian model calculations indicate that an O₃ loss of more than 20 p.p.b.v. per day could have been sustained throughout much of February⁴⁷. Solar radiation and measured ClO increased substantially during the mission⁴⁸, so losses should be much less in early January. Therefore, the loss indicated in our analysis for early January is too large to match these model studies. But meteorological data show that temperatures were low enough to form polar stratospheric clouds at 30 mbar by late November (A. F. Tuck, personal communication), and the eccentricity of the Arctic vortex is sufficient to frequently carry outer vortex air into sunlit regions. These episodic conditions may be sufficient to produce a significant O₃ loss at 30 mbar during late November and December.

Diabatic cooling could bring this somewhat depleted air to ER-2 flight altitudes by early January. Summertime O₃ loss might also have some residual effect.

One of the striking results in the Antarctic campaign was the observed rapid decrease in O₃ and simultaneous increase in ClO as the aircraft crossed into the ozone hole along an isentropic flight leg¹⁰. On 9 February 1989, a similar but less dramatic isentropic O₃ decrease was observed poleward of the Arctic boundary⁴⁹ that was coincident with very high ClO and is shown in Fig. 4a, b. Figure 4a shows decreasing O₃ poleward of the vortex boundary that is collocated with decreasing N₂O. This positive correlation between O₃ and N₂O inside is contrasted by the usual negative correlation seen outside the boundary, implying O₃ loss within the vortex. Figure 4b shows no O₃ loss outside the boundary, substantial loss inside (implied by negative values for O₃ - O₃^{*}), and that the loss occurs precisely where ClO increases dramatically. Figure 4c demonstrates that this correlation between O₃ loss and ClO in the outer vortex persists throughout the mission and is evident even with constant N₂O. In this plot of all outer vortex data, ClO is the abscissa and O₃ - O₃^{*} the ordinate with Θ restricted to 435 ± 15 K. The data are tightly binned by N₂O value and each bin is provided with a linear fit. It is evident that O₃ loss and ClO are positively correlated over a wide range of N₂O, and the linear fits are surprisingly similar, although their slopes increase at the higher N₂O values.

Summary

Evaluating O₃ loss in the Arctic winter is complex because of our limited understanding of the effects of polar dynamics. Although there seems to be a consensus on significant diabatic cooling at high latitudes, and poleward flux into the Arctic vortex, there are uncertainties in the cooling rates and differing opinions on the altitude at which air is supplied to the vortex. Descent and poleward flux into the vortex (regardless of the altitude of the flux), along with concurrent O₃ loss together form a 'flowing processor', accepting air rich in O₃ from lower latitudes, and liberating air somewhat depleted in O₃ through the bottom of the vortex. The relative rates of O₃ loss and its resupply

determine whether a signature of that loss will be evident in isentropic and column analyses. To avoid these problems we have used a chemically conserved tracer, N₂O, to infer Arctic O₃ loss, a method similar to that used to infer loss in the exterior of the Antarctic ozone hole²⁷.

The analysis reveals an *in situ* O₃ loss of 500–1,500 p.p.b.v. (12–35%) throughout the vortex relative to its exterior. This corresponds to a decrease in column O₃ of about 6% over the 4 km depth covered in this analysis. If this loss is of anthropogenic origin emerging during the past decade, the decrease we report is consistent with the 4–10% long-term seasonal column change found by the Ozone Trends Panel²⁵. The rate of O₃ loss has been approximated by assuming vertical descent and a conservative cooling rate. Losses of 9–24 p.p.b.v. per day are obtained for parcels with N₂O from 210 p.p.b.v. to 130 p.p.b.v., respectively. These rates are comparable to but slightly higher than model calculations. Larger cooling rates and deviations from vertical descent will both increase this difference.

An O₃ loss was observed in early January that is too large to match model studies. The Arctic vortex polar asymmetry contrasts with the more symmetric Antarctic vortex, and the asymmetry could enhance early O₃ loss in the Arctic relative to the seasonally comparable period in the Antarctic. This would be due to episodic exposure to sunlight during the low-latitude excursions of the polar jet. The low values of O₃ found on 3 January indicate substantial O₃ loss before the mission and may result from this intermittent exposure.

We have found that the O₃ loss is correlated with elevated ClO in the outer vortex, along with enhancement of O₃ loss where there is maximum exposure of the vortex to sunlight. This suggests, but does not prove, that chlorine is an important factor in the ozone decrease. The agreement between this analysis and model studies is not perfect, particularly in early January. Perhaps alternative explanations of the data will emerge that do not require such large wintertime O₃ loss. In any case, it is clear that accurate evaluation of Arctic O₃ loss must account for the effects of polar dynamics. □

Received 21 December 1989; accepted 19 July 1990.

- Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
- Hofmann, D. J., Harder, J. W., Rolf, S. R. & Rosen, J. R. *Nature* **326**, 59–62 (1987).
- de Zafra, R. L. et al. *Nature* **328**, 408–411 (1987).
- Solomon, S., Mount, G. H., Sanders, R. W. & Schmeltekopf, A. L. *J. geophys. Res.* **92**, 8329–8338 (1987).
- J. geophys. Res.* **94**, Nos. D9 and D14 (1989).
- Proffitt, M. H. et al. *J. geophys. Res.* **94**, 16547–16555 (1989).
- Loewenstein, M., Podolske, J. R., Chan, K. R. & Strahan, S. E. *J. geophys. Res.* **94**, 11589–11598 (1989).
- Brune, W. H., Anderson, J. G. & Chan, K. R. *J. geophys. Res.* **94**, 16649–16663 (1989).
- Chan, K. R., Scott, S. G., Bui, T. P., Bowen, S. W. & Day, J. J. *J. geophys. Res.* **94**, 11573–11587 (1989).
- Proffitt et al. *J. geophys. Res.* **94**, 11437–11448 (1989).
- Anderson, J. G. et al. *J. geophys. Res.* **94**, 11480–11520 (1989).
- Jones, R. L. et al. *J. geophys. Res.* **94**, 11529–11558 (1989).
- Rodriguez, J. M. et al. *J. geophys. Res.* **94**, 16683–16703 (1989).
- Farman, J. C. *Phil. Trans. R. Soc. B* **279**, 261–271 (1977).
- Juckes, M. N. & McIntyre, M. E. *Nature* **328**, 590–596 (1987).
- Hartmann, D. L. et al. *J. geophys. Res.* **94**, 16779–16795 (1989).
- Danielsen, E. F. & Houben, H. *Anthropogene Beeinflussung der Ozonschicht*, 191–242 (DEHEMA-Frankfurt am Main, 1988).
- Tuck, A. F. *J. geophys. Res.* **94**, 11687–11737 (1989).
- Proffitt, M. H. et al. *J. geophys. Res.* **94**, 16797–16813 (1989).
- Murphy, D. M. et al. *J. geophys. Res.* **94**, 11669–11685 (1989).
- Watterson, I. G. & Tuck, A. F. *J. geophys. Res.* **94**, 16511–16525 (1989).
- Cariolle, D., Lasserre-Bigorry, A. & Royer, J. F. *J. geophys. Res.* **95**, 1883–1898 (1990).
- Hofmann, D. J. et al. *Nature* **340**, 117–121 (1989).
- Evans, W. F. *J. Geophys. Res. Lett.* **17**, 167–170 (1990).
- Ozone Trends Panel *Present State of Knowledge of the Upper Atmosphere, 1988*, NASA Ref. Publ. 1208 (Natn. Tech. Inf. Serv., Springfield, Virginia, 1989).
- Geophys. Res. Lett.* **17**, No. 4 (1990).
- Proffitt, M. H., Fahey, D. W., Kelly, K. K. & Tuck, A. F. *Nature* **342**, 233–237 (1989).
- Solomon, S., Garcia, R. R. & Stordal, F. *J. geophys. Res.* **90**, 12981–12989 (1985).
- Garcia, R. R. & Solomon, S. *J. geophys. Res.* **90**, 3850–3868 (1985).
- Brasseur, G. J. & Solomon, S. *Aeronomy of the Middle Atmosphere* 2nd Edn (Reidel, Dordrecht, 1986).
- Perilski, L. M., Solomon, S. & London, J. *Planet. Space Sci.* **37**, 1527–1538 (1989).
- Strahan, S. E. et al. *J. geophys. Res.* **94**, 16749–16756 (1989).
- Gunson, M. R. et al. *J. geophys. Res.* (in the press).
- Rosenfeld, J. E., Schoeberl, M. R., Lait, L. R. & Newman, P. A. *Geophys. Res. Lett.* **17**, 345–348 (1990).
- Schoeberl, M. R. et al. *Geophys. Res. Lett.* **17**, 469–472 (1990).
- Brewer, A. W. *Q. J. R. met. Soc.* **75**, 351–363 (1949).
- Dobson, G. M. G. *Proc. R. Soc. A* **236**, 187–193 (1956).
- Brewer, A. W. & Wilson, A. W. *Q. J. R. met. Soc.* **94**, 249–265 (1968).
- Dütsch, H. U. *Can. J. Chem.* **52**, 1491–1504 (1974).
- London, J. F., Bojkov, S., Oltmans, S. & Kelley, J. I. *Atlas of the Global Distribution of Total Ozone July 1957–June 1967* NCAR tech. Note 113 (NCAR, Boulder, 1976).
- Bojkov, R. D. & Rumen, D. *Met. atmos. Phys.* **38**, 117–130 (1988).
- Ozone Data for the World, Atmospheric Environment Service* Vol. No. 3 30, 312–318 (Can. Atmos. Envir. Serv. & World Met. Org., Toronto, 1989).
- Leovy, C. B. et al. *J. atmos. Sci.* **42**, 230–244 (1985).
- Kent, G. S., Trepte, C. R., Farrukh, U. H. & McCormick, M. P. *J. atmos. Sci.* **42**, 1536–1551 (1985).
- Farman, J. C., Murgatroyd, R. J., Silnickas, A. M. & Thrush, B. A. *Q. J. R. met. Soc.* **111**, 1013–1028 (1985).
- Salawitch, R. J. et al. *Geophys. Res. Lett.* **17**, 561–564 (1990).
- McKenna, D. S. et al. *Geophys. Res. Lett.* **17**, 553–556 (1990).
- Brune, W. H., Toohay, D. W., Anderson, J. G. & Chan, K. R. *Geophys. Res. Lett.* **17**, 505–508 (1990).
- Browell, E. V. et al. *Geophys. Res. Lett.* **17**, 325–328 (1990).

ACKNOWLEDGEMENTS. We thank Ed Danielsen, Stuart McKeen, George Mount, Susan Solomon, Michael Trainer and Adrian Tuck for their helpful comments, William Brune and Darin Toohay for the use of the ClO data, Melanie Steinkamp and Cynthia Proffitt for their help in Stavanger, Ken Aikin for producing the figures, and the ER-2 pilots and ground crew.

Atomic structure of the actin:DNase I complex

Wolfgang Kabsch, Hans Georg Mannherz*, Dietrich Suck†, Emil F. Pai & Kenneth C. Holmes

Max-Planck-Institut für medizinische Forschung, Abteilung Biophysik, Jahnstraße 29, 6900 Heidelberg, FRG

* Institut für Anatomie und Zellbiologie der Philipps-Universität, Robert-Koch-Straße 6, 3550 Marburg, FRG

† European Molecular Biology Laboratory, Postfach 102209, 6900 Heidelberg, FRG

The atomic models of the complex between rabbit skeletal muscle actin and bovine pancreatic deoxyribonuclease I both in the ATP and ADP forms have been determined by X-ray analysis at an effective resolution of 2.8 Å and 3 Å, respectively. The two structures are very similar. The actin molecule consists of two domains which can be further subdivided into two subdomains. ADP or ATP is located in the cleft between the domains with a calcium ion bound to the β - or β - and γ -phosphates, respectively. The motif of a five-stranded β sheet consisting of a β meander and a right handed $\beta\alpha\beta$ unit appears in each domain suggesting that gene duplication might have occurred. These sheets have the same topology as that found in hexokinase.

ACTIN was discovered by Straub^{1,2} in an extract of muscle tissue. Together with myosin it forms the contractile protein complex, actomyosin. In the absence of salt it can exist as a globular protein, G-actin. On addition of traces of salt it changes into a highly asymmetrical fibrous protein, F-actin. Ever since its discovery, actin has been the subject of intensive studies in many laboratories because of its central role in muscle and cell motility. Some topics of current research have been reviewed by Pollard³ and Vandekerckhove⁴.

As normally isolated, G-actin from rabbit skeletal muscle consists of a single polypeptide chain, a nucleotide (ATP) and a divalent cation bound to a specific site. The complete amino-acid sequence has been determined⁵⁻⁸. Among the 375 residues are 5 free sulphhydryl residues, at positions 10, 217, 257, 285, and 374. Acetylation of the N-terminal aspartate and methylation of histidine 73 are post-translational modifications.

The tendency of actin to polymerize prevents the formation of crystals—a necessary requirement for determining the three-dimensional structure by X-ray analysis. The capability of actin to polymerize is lost, however, after forming a 1:1 complex with bovine pancreatic deoxyribonuclease I (DNase I)⁹⁻¹¹. This complex can be crystallized^{12,13} and its structure has been determined at low resolution^{14,15}. More recently, the structure of DNase I without actin has been solved at atomic detail^{16,17}. We now report the three-dimensional structure of the actin:DNase I complex at 2.8 Å resolution derived from knowledge of the DNase I coordinates and two heavy-atom derivatives.

Structure determination

The classic method of data acquisition on films requires many crystals for a complete data set because of the long exposure times necessary to collect high resolution data. Despite considerable effort we were unable to extend the resolution of the actin:DNase I structure beyond 4.5 Å (ref. 15) because of the variability of the unit cell parameters and problems of reproducibility of the heavy-atom derivatives. This situation has been radically improved by the development and commercial availability of

electronic area detectors. We have designed a program^{18,19} to extract diffraction data of good quality from the three-dimensional reflection profiles recorded by an electronic area detector. As shown in Table 1 almost complete heavy-atom derivative data to 3.5 Å resolution have been collected, each from only one crystal. The native data at higher resolution are from a selected set from several crystals. A summary of the structural analysis by the isomorphous replacement method based on these data is given in Table 1a. Portions of the resulting electron density map at a nominal resolution of 3.5 Å were shown in an earlier paper²⁰. Although long stretches of the polypeptide chain can be followed in this map many ambiguous regions prevented a direct fit of an atomic model to the electron density derived solely from double isomorphous replacement and solvent flattening. Additional information provided by the known structure of DNase I representing about 40% of the atoms in the unit cell had to be exploited. As the method we have used for the successful solution is likely to be useful for other difficult structure determinations, it is described here.

A cyclic procedure was developed which exploits the knowledge of the DNase I atomic model to phase native reflections to higher resolution and to improve phases obtained from the heavy-atom derivatives at lower resolution. Each cycle of this procedure was carried out in the following way: The current atomic model was refined with the program XPLOR²¹ using the standard protocol for heating, fast cooling and minimization. The electrostatic interaction energy was reduced by a factor of four. The trajectory was followed for 2 ps at 2,000 K and for 0.5 ps in the cooling stage. A fixed temperature factor of 20 was assigned to all atoms of the model. After the final minimization, restrained isotropic temperature factors were determined for each atom. Wrongly placed atoms would thus be weighted down as they receive high temperature factors. Phase information from the refined atomic model and from the heavy-atom derivatives was then combined in a way similar to the method of Read²² which reduces model bias from the new electron density map. The improved map is then used to correct errors and to extend the partial atomic model. Atoms were fitted to the density using the interactive graphics program FRODO²³ adapted to an IRIS 4GT (Silicon Graphics Inc., Mountain View, California) by C.M. Cambillau.

The cyclic procedure just described was started with the atomic model of DNase I comprising about 40% of the atoms in the crystal. After the first cycle a map of the actin density was plotted on plexiglass sheets. An overview of the complete electron density map was found necessary for developing ideas about the overall chain fold in the context of the known properties of actin. Map interpretation was started at the mercury position labelling Cys 10 in actin. This heavy-atom position is halfway between a β sheet and a bulky helix in the small domain of actin. Considering the whole map of actin it seemed most likely that a unique stretch of large side chains in the sequence between residues 79–91 should be assigned to this bulky helix. As a consequence Cys 10 should then belong to the β sheet. In addition, evidence from crosslinking experiments and enzymatic cleavage studies indicates that some residues within the region 50–69 of the actin sequence should be involved in binding to DNase I. Once interpretable stretches of the actin chain had been identified, C $_{\alpha}$ - and some C $_{\beta}$ -positions of these residues

were marked in the plexiglass map. These positions were used to generate a complete atomic model by the program XCHAIN (W.K., unpublished results). The program was designed to search the data bank of known structures for pentapeptides with the central residue identical to each actin residue in succession. A complete atomic model was assembled from the central residues of pentapeptides best fitting the marked positions in the plexiglass map as well as minimizing the discrepancies between main chain atoms in the overlap region of the already selected pentapeptides. Typically about 40–70 new actin residues were included in each new step of the cyclic procedure described above. About 20 such cycles were carried out, as many corrections to the model had to be done in the difficult regions of the map.

At present the structure has been refined to *R*-factors below 20% (see Table 1a). The root mean square (r.m.s.) deviation of the model from ideal bond length, bond angles and peptide planarity are 0.017 Å, 3.3° and 2.0°, respectively. About 22 ϕ/ψ -angles of the actin polypeptide chain are outside the allowed region in the Ramachandran plot. The structure of ATP-actin:DNase I was refined to similar values using the ADP-actin:DNase I structure as a starting model. Atomic coordinates will be deposited with the Brookhaven Protein Data Bank (see Supplementary Information).

Backbone structure

The course of the polypeptide chains of the ATP-actin:DNase I complex is shown in Fig. 1a. The viewing direction is roughly

FIG. 1 Structure of ATP-actin:DNase I. *a*, C_{α} -stereo plot. Open circles mark C_{α} -positions of actin residues. The C_{α} -atoms of DNase I are connected by thin lines. Residues 102 and 103 are omitted as their positions have not been assigned. Three Ca^{2+} in the DNase I region and a single Ca^{2+} near the phosphates of ATP in actin are shown by filled circles. *b*, Schematic representation²⁴ of the three-dimensional structure of actin shown in the same orientation as *a*. First and last amino-acid residues in the helices and sheet strands are shown by filled circles. The assignment of secondary structure is based on the automatic procedure of Kabsch and Sander²⁵. However, in the drawing some of the helices and sheet strands have been extended by one or two residues beyond the strict assignments where geometry indicates that the secondary structure was likely to become more extended when refinement is complete.

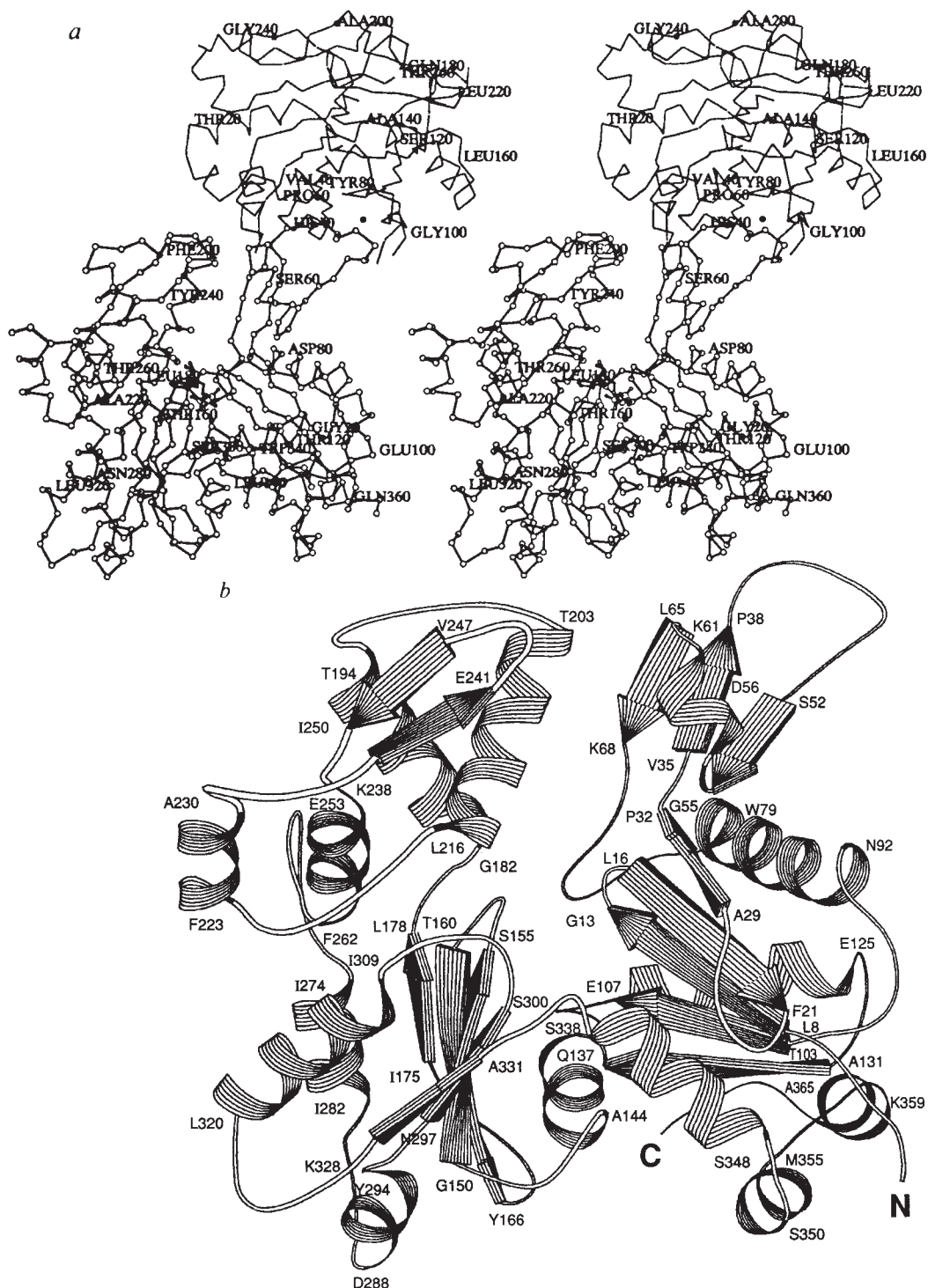


TABLE 1 Statistics of crystallographic data

a ADP-actin:DNase I (cell parameters a=132.1 Å, b=56.6 Å, c=110.1 Å)						
Native	Overall	∞-6.0	Resolution shells (Å)			
			6.0-4.5	4.5-3.5	3.5-2.8	2.8-2.4
Number of measured reflections (2 crystals)	63,259	11,438	10,850	19,243	21,728	
Number of measured unique reflections	17,631	2,203	2,891	5,379	7,158	
Number of possible unique reflections	21,030	2,293	2,974	5,667	10,096	
R_f (%) [*]	5.7	4.7	4.8	6.0	11.6	
Completeness (%) of						
data with $1/\sigma > 1$	79	94	95	91	63	
data with $1/\sigma > 3$	68	91	91	84	46	
R (%) [†] using						
data with $1/\sigma > 1$	18.8	23.0	16.5	16.0	23.8	
data with $1/\sigma > 3$	17.5	22.4	16.0	15.2	21.2	
Mean figure of merit	0.60	0.83 [‡]	0.65	0.47		
Pb(NO ₃) ₂ (a=132.8 Å, b=56.7 Å, c=110.3 Å)						
Number of measured unique reflections	10,306	2,185	2,829	5,292		
Phasing power [§]	1.6	1.8	1.8	1.4		
R_F [%] [¶]	17.8	17.5	17.7	18.0		
R_C [%]	58.9	49.3	55.3	71.1		
CH ₃ HgOAc (a=132.5 Å, b=56.5 Å, c=110.1 Å)						
Number of measured unique reflections	10,668	2,156	2,935	5,577		
Phasing power [§]	1.4	1.6	1.5	1.2		
R_F [%] [¶]	21.4	19.9	19.8	23.1		
R_C [%]	67.1	57.5	64.6	77.8		
b ATP-actin:DNase I (cell parameters a=132.9 Å, b=56.3 Å, c=109.7 Å)						
Number of measured reflections (4 crystals)	103,367	8,975	12,262	21,269	31,401	29,460
Number of measured unique reflections	31,362	2,182	2,898	5,557	9,854	10,871
Number of possible unique reflections	32,996	2,288	2,965	5,641	10,064	12,038
R_f (%) [*]	10.7	6.1	6.7	8.5	18.6	43.1
Completeness (%) of						
data with $1/\sigma > 1$	81	94	96	95	87	65
data with $1/\sigma > 3$	48	89	90	82	49	12
R (%) [†] using						
data with $1/\sigma > 1$	20.8	25.4	17.4	16.5	22.5	27.0
data with $1/\sigma > 3$	17.6	24.3	16.5	15.4	19.2	19.8

Orthorhombic crystals (space group $P2_12_12_1$) of the actin:DNase I complex were obtained at 4 °C at pH 6.6 by adding 8–10% polyethyleneglycol 6000 as precipitating agent to a solution containing about 5 mg ml⁻¹ protein, 50 mM imidazole, 0.1 mM CaCl₂, 1 mM NaN₃ and 2 mM MgATP¹⁴. Before crystallization the complex was subjected in solution to a mild tryptic treatment which removes three amino acids—including Cys 374—at the C terminus of actin. Data of ADP-actin:DNase I were obtained from older crystals as ATP slowly hydrolyses into ADP. The ATP-actin:DNase I data were collected from crystals grown within 2–4 weeks. X rays (CuK α) were generated by a GX-18 rotating anode (Elliot/Enraf-Nonius, Delft) and focussed by Franks double-mirror optics. The anode was operated at 35 kV and 50 mA. An oscillation camera (Supper, Natick, MA, USA) with a vertical rotation axis was used to move the crystal. Diffraction data were recorded by an electronic area detector (Siemens/Nicolet) at a distance of 13 cm from the crystal. Exposure time was 3–4 min for each data frame covering a rotation range of 0.1667° about the spindle axis. Crystals were cooled to about 4 °C during the measurement. Camera and detector were controlled by a simplified version of a program by Blum *et al.*⁶² running on a small dedicated computer (PCS, München, FRG). For each data collection run a sequence of adjacent rotation pictures was transferred via Ethernet to a large computer (CONVEX C210) and processed by the program XDS¹⁸ which includes routines for the automatic indexing of the observed diffraction pattern¹⁹. The program terminates with a list of squared structure factor amplitudes corrected for background, absorption and radiation damage. The Pb(NO₃)₂ and CH₃HgOAc derivatives were obtained by soaking crystals of ADP-actin:DNase I in 1 mM heavy-atom solutions for 1 day and 1 h, respectively.

^{*} $R_f = \sum_i |I_{hi} - I_h| / \sum_i I_{hi}$ where h are unique reflection indices and I_{hi} are the intensities of symmetry equivalent reflections giving a mean value of I_h .

[†] $R = \sum |F_{obs} - F_{model}| / \sum F_{obs}$ where F_{obs} and F_{model} are observed and atomic model structure factor amplitudes, respectively. Reflections at a lower resolution than 10 Å are excluded from the summation.

[‡] Additional heavy-atom derivatives from earlier work are included.

[§] Phasing power is the mean value of the heavy-atom structure factor amplitude divided by the residual lack-of-closure error.

[¶] $\sum |F_{derivative} - F_{native}| / \sum F_{native}$.

^{||} Cullis R -factor for centric reflections.

perpendicular to the flat face of actin. Elements of secondary structure of actin are specified in Fig. 1b. The molecule as drawn fits into a square of sidelength 55 Å. The flat side is of width 35 Å. Actin consists of two domains (for historical reasons called large and small, although they turn out to be of almost the same size) with the nucleotide (ATP or ADP) and an associated calcium ion bound in the cleft between the two domains. Both N and C termini are in the small domain. Each of the domains consists of two subdomains. By definition, subdomains 1 (residues 1–32, 70–144, and 338–372) and 2 (residues 33–69) constitute the small domain whereas subdomains 3 (residues 145–180, and 270–337) and 4 (residues 181–269) represent the large domain of actin.

Subdomain 1 contains a five-stranded β -pleated sheet

assembled from a β meander and a right-handed $\beta\alpha\beta$ unit of the same topology as that one found in hexokinase²⁴. The sheet is surrounded by five helices. One of these helices contains two hydrophobic residues, Ile 345 and Leu 346, which are accessible to solvent. Their exposed areas²⁵ are about 65 Å² and 32 Å², respectively. In contrast the preceding hydrophilic residue Ser 344 is completely buried. Its side chain forms a hydrogen bond with the main chain carbonyl group of Asp 24. An error in assignment is unlikely as all residues of this helix—including Trp 340, which is completely buried—fit the electron density map very well. We speculate that this helix can be involved in interactions with an actin-binding protein, possibly myosin or tropomyosin.

Subdomain 2 consists of a three-stranded antiparallel β -

pleated sheet with a helix connecting the two edge-strands. Parts of the loop connecting the first with the second strand are attached as an additional strand to a β sheet in DNase I.

Subdomain 3 contains a five-stranded β sheet which is surrounded by three helices. The sheet topology is identical to that in subdomain 1. Moreover, the α -carbon positions of 75 pairs of corresponding residues can be superimposed by an approximate twofold rotation with a r.m.s. error of 2.8 Å. The rotation angle is 168° and the component of the translation vector along the rotation axis is 0.6 Å. This suggests that actin has evolved by gene duplication although the amino acid sequence does not display any significant internal similarity. It is possible that soon after the duplication event, each of the domains was optimized

to carry out quite different functions. The idea that both domains of actin evolved from a common ancestral protein agrees well with the identification of a hybrid molecule containing residues 1–134 of actin and a tyrosine-specific protein kinase²⁶.

Subdomain 4 contains a two-stranded antiparallel β sheet and four α helices. Contrary to normal definitions, residues Glu 253 to Phe 262 are shown in the helical conformation in Fig. 1b although the characteristic hydrogen bond pattern is disrupted by Pro 258. The distance between the sulphur atoms of Cys 217 and Cys 257 is about 3.8 Å and both residues are inaccessible to solvent. This helix is followed by a 10-residue loop and another helix (residues 274–282) in subdomain 3. Both helices can be considered as parts of a single large helix of length ~40 Å interrupted by a loop in the middle. Three hydrophobic residues in this loop, Phe 266, Ile 267, and Met 269 are accessible to solvent. Their exposed area²⁵ is about 60 Å² each. We believe that this loop is involved in a contact across the helix axis with another pair of actin molecules in the F-actin filament (see accompanying paper²⁷).

Individual isotropic temperature factors were determined for each atom. Between bonded atoms these values were restrained to reasonably small differences. Regions in the actin model with temperature factors above 40 are: 1–7, 95–100, 183–276, 306–329 and 348–373. It has not yet been established whether the mobility of the structure results from a collective movement of large parts—like subdomain 4—or from uncorrelated thermal motion. Even in regions of high temperature factors the atomic model is unique except for possible local errors and agrees well with the electron density. Uncertain regions of our model are the actin residues 1–6, 232–235 and 366–373. Actin appears as a rather flexible structure. Its four subdomains are held together mainly by the nucleotide, and by salt bridges. The complex formation with DNase I involving subdomains 2 and 4 adds further to the stability of the structure. This is in harmony with the increased stability of nucleotide-free actin against denaturation when complexed to DNase I²⁸.

DNase I binding site

The major contact between actin and DNase I consists of hydrogen bonds, electrostatic and hydrophobic interactions as shown in Fig. 2. Residues Gly 42, Val 43 and Met 44 are incorporated as a parallel strand of a β sheet in DNase I with partners Tyr 65, Val 66, Val 67. The minor contact consists of electrostatic interactions between residues Thr 203 and Glu 207 in subdomain 4 and Glu 13 and His 44 in DNase I. The atomic model of the

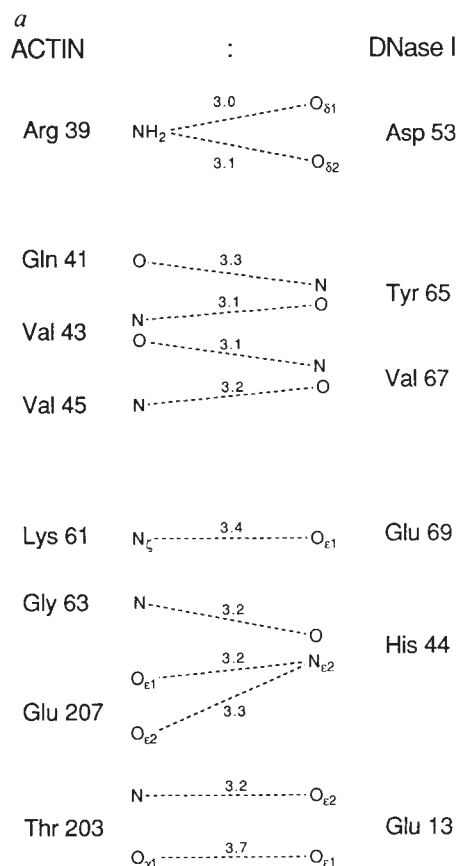


FIG. 2 Binding regions between actin and DNase I. *a*, Atoms involved in hydrogen bonds and electrostatic interactions between the two proteins. Distances are given in Å. The actin residues Gly 42, Val 43 and Met 44 are incorporated as a parallel strand of a β -pleated sheet in DNase I with partners Tyr 65, Val 66, and Val 67. *b*, Stereo-plot of all atoms involved in the binding regions. The backbone atoms in actin are connected by thicker lines than those in DNase I.

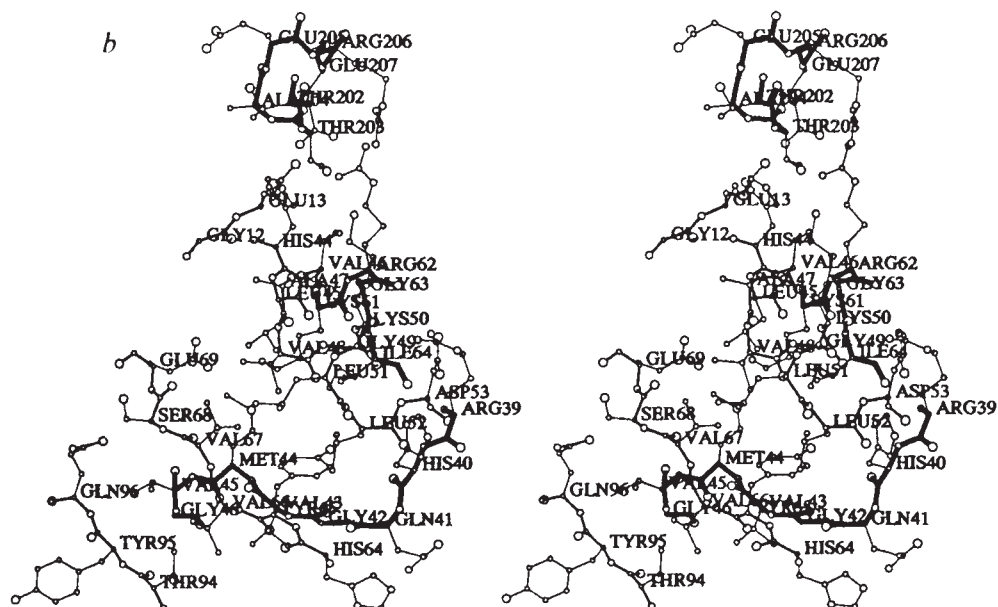


TABLE 2 Distances between fluorescence-labelled positions

	Cys 10	Tyr 69	Cys 374	Metal	Adenine	Ribose
Cys 10		19	17	12	22	19
Tyr 69			29	16	18	11
Cys 374	(32)	(39)		17	30	28
Metal	(30)	(21)			14	11
Adenine	(40)	(21)	(33)	(16) (<10)		8
Ribose			(34)			

Mutual distances are calculated between the oxygen position of the hydroxyl group of Tyr 69, the S_γ-positions of Cys 10 and Cys 374, the calcium position, the N6-atom of the adenine ring, and the centroid of the O2'- and O3'-atoms of the ribose. The location of Cys 374 was estimated by model building as the three C-terminal residues are missing in our structure. Distances obtained from fluorescence resonance energy transfer measurements (reviewed in ref. 63) are shown in brackets. All distances are in Ångströms. With the exception of the distances involving Cys 10, the results from the fluorescence resonance energy transfer measurements agree with our structure. It seems likely that treatment of actin by 1–2 M urea to enable Cys 10 to be labelled has greatly altered the structure.

actin-DNase I interaction explains a large number of observations. Inhibition of DNase I activity by actin⁹ occurs by steric hindrance as Glu 13 of DNase I is also involved in binding of DNA²⁹. The complex cannot be dissociated by the addition of salt because of the hydrophobic packing between residues of both proteins. Immunological studies³⁰ show that the binding of DNase I to actin is affected by antibodies specific to a central region between residues 168–226 in the actin sequence. In addition, the N- and C-terminal regions of actin were found to be in spatial proximity at the surface of the actin monomer and

far from the binding region of DNase I. Crosslinking experiments of the actin:DNase I complex show that one of the lysine or tyrosine residues within the region 50–69 of the actin sequence must be involved in binding to DNase I³¹. In our model we find that Lys 61 forms a salt bridge with Glu 69 in DNase I. Enzymatic cleavage of actin by several proteases of varying specificity has revealed a number of proteolytic sites near residues 29, 40, 44, 47, 61, 68, 114, 207, 226, and 372 (refs 32–37) all of which were found at exposed regions in our model. It was also observed that proteolysis near residues 61–68 does not occur for the actin:DNase I complex. Subtilisin cleavage between Met 47 and Gly 48 greatly reduces the binding between actin and DNase I³⁷.

The binding is very tight, causing a distortion of the DNase I structure from its native form by as much as 1.8 Å in the C_α-positions of residues 65–67. It is likely then that the actin structure is also disturbed in the regions of subdomain 2 which are in close contact with DNase I. X-ray analysis of crystals of the actin:profilin complex is in progress^{38,39}, and this is expected to provide an undistorted structure of subdomain 2 as the contact region between these two proteins is near Glu 364 in subdomain 1⁴⁰.

Nucleotide-binding site

The adenine nucleotide is bound in the cleft between the small and large domain. As DNase I bridges subdomains 2 and 4, it is understandable that the nucleotide exchange rate is reduced on complex formation²⁸. Structural details of the interaction of ADP and ATP with actin are shown in Figs 3–5.

The adenine base fits into a pocket formed by residues Lys 213, Glu 214, Thr 303, Met 305, Tyr 306, and Lys 336. There are no specific interactions between these amino acids and the adenine

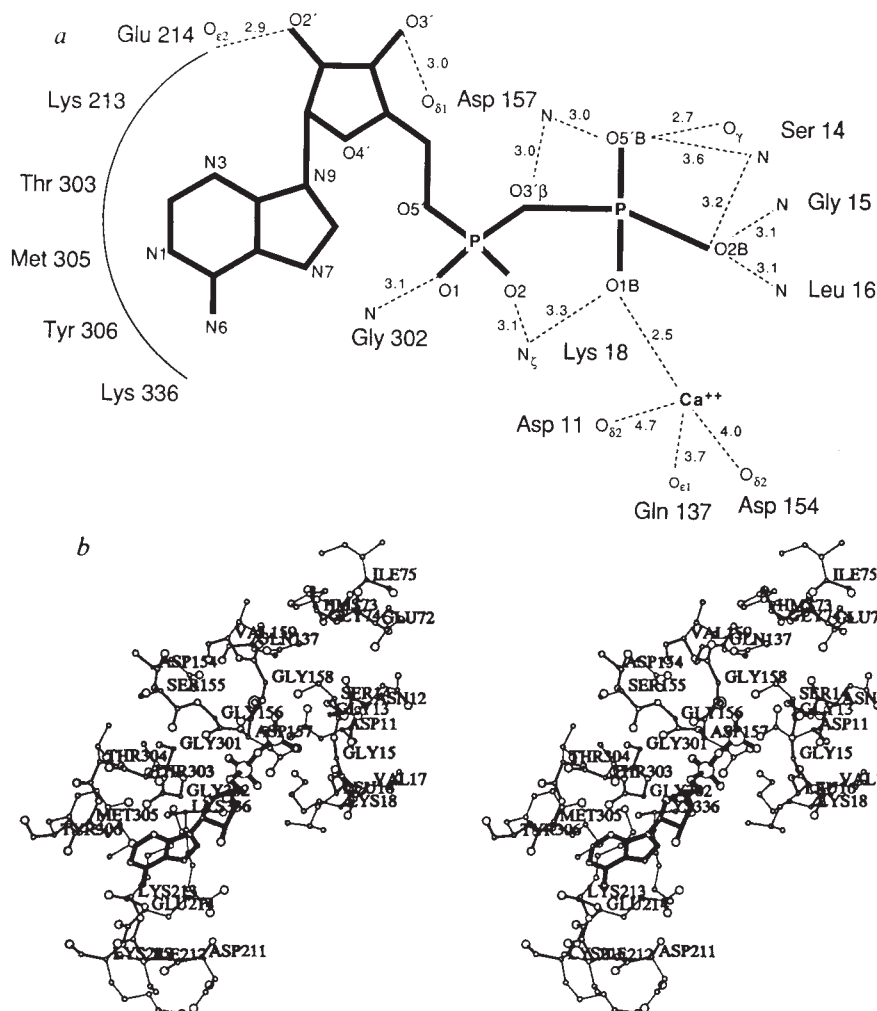


FIG. 3 Environment of ADP in actin. *a*, Schematic drawing of nucleotide interactions with actin and calcium. Distances between atoms are given in Å. Amino acids of actin specified near the circular segment in the left part of the drawing form a pocket covering the adenine. *b*, Stereo-plot of the atoms in the environment of ADP.

base. Tyr 306 and Lys 336 had been identified earlier to be close to the adenine base. Tyr 306 can be labelled by 2-azido-ADP⁴¹. For the nucleotide analogue 8-azido-ATP the label was bound primarily to Lys 336 both in F- and G-actin⁴². In G-actin this derivative reacts also with Trp 356, though to a lesser extent. We have no explanation for this observation in terms of our model. Trp 356 is buried in the structure at a distance of about 27 Å from the 8-position of the base. Perhaps actin is in a partially denatured state when the label reacts with Trp 356.

The ribose ring is in the 2'-endo conformation. The dihedral angle χ of the *N*-glycosidic bond is -125° which is within the anti-range observed for nucleotides with 2'-endo puckering of the ribose ring⁴³. The 2'- and 3'-hydroxyl groups of the ribose participate in hydrogen bonds with one of the oxygen atoms of the carboxylate groups of Glu 214 and Asp 157, respectively.

The phosphate groups of ADP (Fig. 3) and ATP (Fig. 4) are involved in a large number of interactions. The O1-atom of the α -phosphate and the main chain amide group of Gly 302 form hydrogen bonds in both the ADP- and ATP-structures. In the case of ADP, the ϵ -amino group of Lys 18 forms a bifurcated hydrogen bond with the O2-atom of the α -phosphate and the O1-atom of the β -phosphate. In the ATP-structure, Lys 18 interacts with the main chain carbonyl and the side chain carboxylate group of Asp 11, instead. The β - and γ -phosphates are involved in hydrogen bonds with amides of residues Ser 14, Gly 15, Leu 16 and residues Asp 157, Gly 158, Val 159, respectively. These residues belong to loops connecting the first with the second strand in each of the homologous β sheets in subdomains 1 and 3. As shown in Fig. 5 an approximate two-fold axis relates

the β - and γ -phosphates, the two β bends, and some side chains of the β bends, in particular the corresponding Asp residues 11 and 154. The symmetry is not perfect, however: whereas the γ -phosphate oxygen O1C is equally hydrogen bonded by the three amide groups from Asp 157, Gly 158, and Val 159, the β -phosphate oxygen O2B is only hydrogen bonded to Gly 15 and Leu 16, the distance from the amide group of Ser 14 (3.9 Å) is too long to be a hydrogen bond. This might indicate a small distortion to the structure introduced by formation of the tight complex with DNase I. In the ADP-form the β -phosphate oxygen O2B forms all three hydrogen bonds and sits in the centre of the Ser 14, Gly 15, Leu 16 bend. The hydrogen bonds may be of significance for the mechanism of the ATPase where the tight coordination by the amide groups will serve to extract electrons from the β -phosphate O2B and the γ -phosphate O1C. The proximity of the Ser 14 O_γ to the γ -phosphate may also be significant.

Calcium-binding site

Four calcium ions have been identified in the electron density map. Three are bound to DNase I. The first two which are near the marked residues Ala 200 and Gly 100 in Fig. 1a also occur in the native DNase I structure without actin. An additional calcium ion is located near Glu 193. The only calcium ion found within the actin molecule sits in a deep hydrophilic pocket formed by the phosphate groups of the adenine nucleotide and the actin residues Asp 11, Gln 137, and Asp 154. These residues seem to have only an indirect role in the coordination of the calcium ion as they are too far away for a direct interaction. In

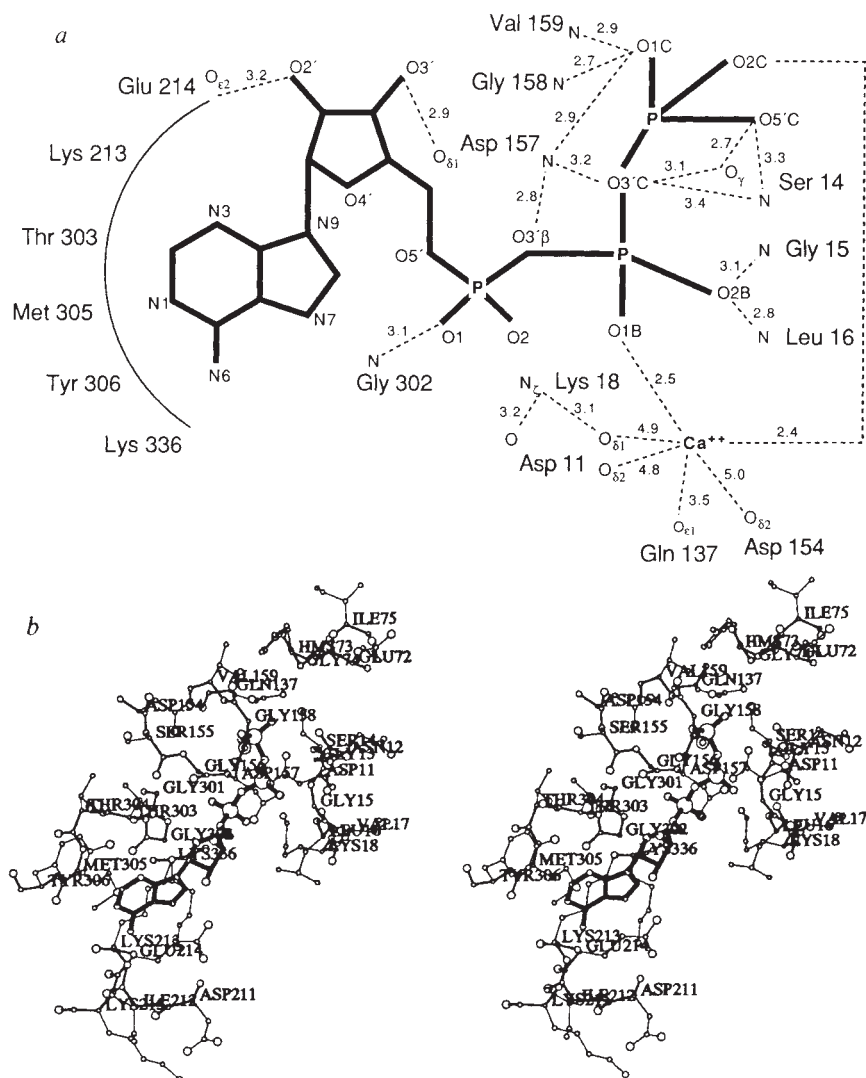


FIG. 4 Environment of ATP in actin. *a*, Schematic drawing of nucleotide interactions with actin and calcium. Distances between atoms are given in Å. Amino acids of actin specified near the circular segment in the left part of the drawing form a pocket covering the adenine. *b*, Stereo-plot of the atoms in the environment of ATP.

the ADP-actin:DNase I structure, calcium is bound to the β -phosphate oxygen at a distance of 2.5 Å (Fig. 3). In the ATP-actin:DNase I structure calcium is tightly bound in the Λ -configuration⁴⁴ to the β - and γ -phosphate oxygens at distances of 2.5 Å and 2.4 Å, respectively (Figs 4 and 5). Using the Cahn-Ingold-Prelog nomenclature, the metal ion is bound to the pro-S oxygen of the β -phosphate group. It is known that actin has a single high-affinity site for divalent cations and several low-affinity sites for mono-, di- or trivalent cations. Occupancy of these sites reduces the net negative charge of the monomers and promotes polymerization. For rabbit skeletal muscle actin there are four low-affinity Ca^{2+} sites, one of which is expected to be near an acidic residue at the N terminus⁴⁵. As the six N-terminal residues of actin are very mobile and the electron density is very low in this region of our map we are unable either to confirm or reject this hypothesis. Biochemical evidence on the location of the high-affinity site has led to contradictory conclusions. It has been suggested that the nucleotide and metal ion binding sites are spatially separated⁴⁶ or associated⁴⁷⁻⁵⁰. As we find only one bound calcium, we believe this to be the high-affinity site. Furthermore, as shown in Table 2, various distances between calcium, ribose, adenine, Tyr 69 and Cys 374 calculated from our structure agree with fluorescence resonance energy transfer measurements. We cannot, however, definitely exclude the possibility that our calcium is at a low-affinity site and that the high-affinity site is occupied by magnesium, which could be difficult to detect in the map.

Myosin-binding sites

The atomic model of ADP-actin has been used successfully to explain the X-ray diffraction data obtained from oriented gels of F-actin²⁷. The unique orientation of the monomer thus found has the small domain at large radius from the F-actin helix axis. All four subdomains of actin are involved in contacts with neighbouring monomers. From three-dimensional image reconstruction of electron micrographs of actin filaments decorated with myosin heavy chain subfragment 1 (S1) it has been shown that the elongated S1 molecule is connected to actin at large radius and points away from the filament axis⁵¹. We expect therefore that myosin binds to the small domain of actin. This view is supported by a large number of observations which led to the identification of amino-acid residues involved in binding. Indeed, all these residues are located in the small domain of actin. Results from crosslinking experiments⁵² indicate that some acidic residues of actin at positions 1-4 and 11 are involved in the binding of S1 in the rigor state. Using a different protein cross-linker, Bertrand *et al.*⁵³ have identified actin peptides 1-28 and 40-113 to participate in forming the rigor complex. Antigenic probes have located the main interaction in the neighbourhood of the actin residues 18-28 (ref. 54) whereas the N-terminal

region 1-7 seems to be near but not directly in contact with the S1 binding site⁵⁵. Results from proton NMR studies⁵⁶ also indicate that the actin residues 21-40 are involved in binding S1. Furthermore, it has been concluded that Cys 10 is close to the myosin-binding site⁵⁷, that Lys 61 and Lys 336 are close to attachment sites for both S1 and tropomyosin^{58,59}. If we locate the residues thought to be involved in S1 binding in the actin structure the following picture emerges: the central contact region consists of residues Asp 24, Asp 26, Arg 28 in the loop connecting the second with the third strand in the sheet of subdomain 1 and extends towards residues Glu 93 and Arg 95 in the loop connecting the bulky helix with the fourth strand in the sheet. These two loop regions are found on the same flat side of actin at a distance of about 12 Å (see Fig. 1). The contact region is delimited by Lys 336, by the acidic residues 1-4 at the N terminus and by Lys 61 on the short helix connecting the second with the third strand in subdomain 2. It is conceivable that Cys 10 feels the binding of myosin as it is buried in the molecule about 10 Å away from Arg 28. As mentioned earlier it is possible that the two hydrophobic residues, Ile 345 and Leu 346, which are accessible to solvent, are involved in binding tropomyosin or myosin as they are at a distance of about 16 Å from both Lys 336 and Arg 28.

It has also been shown^{52,60} that the myosin alkaline light chain 1 binds to the C-terminal residues Glu 361, Asp 363 and Glu 364 of actin. Furthermore, a number of residues in the actin sequence have been pinpointed in the literature to be involved in the binding of phalloidin, actin, tropomyosin, troponin I, profilin, α -actinin, gelsolin, and other molecules interacting with actin.

Implications

The atomic structure of actin as determined from the complex with DNase I is a useful reference frame for organizing the enormous amount of data from chemical and physical studies of this molecule described in the literature. Only a small fraction of these results have been mentioned in this article. As described in the accompanying paper²⁷, we were able to derive the orientation of the actin molecule in the F-actin filament. Knowledge of the atomic model of the actin filament should contribute to be better understanding of the mechanism of muscle and also of the actin cytoskeleton and its control system in non-muscle cells.

Structural analysis of actin has yielded some unexpected results which were not detectable at the amino-acid sequence level. These are the similar architecture of subdomains 1 and 3 and the common features between actin and hexokinase. In addition we find an almost perfect structural agreement between actin and the N-terminal ATPase fragment (relative molecular mass 44,000; M_r 44K) of the 70K bovine heat shock cognate

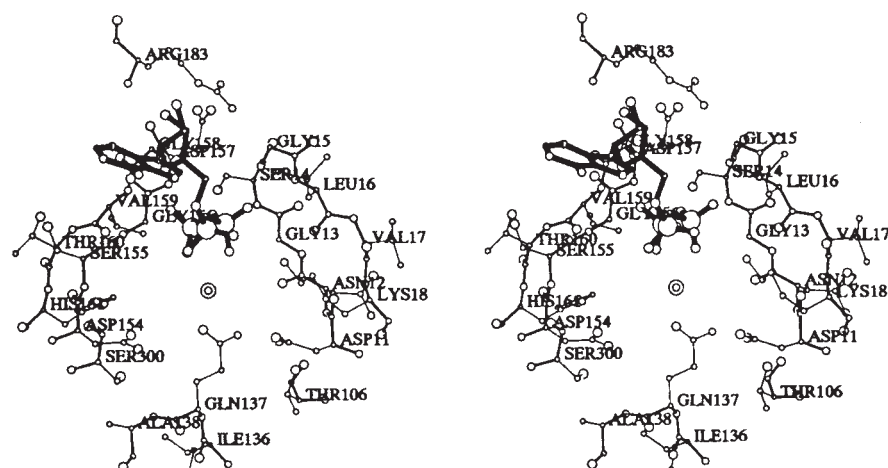


FIG. 5 Stereo-plot of the environment of calcium in ATP-actin:DNase I. The calcium ion is in the centre of the picture. The atoms belonging to ATP are connected by thick lines. A pseudo two-fold axis (roughly vertical in the figure) passes through the calcium ion and relates the two phosphate-binding loops and associated β sheets.

protein (HSC 70). A comparison of actin with the HSC 70 structure as described by Flaherty *et al.*⁶¹ leads to the following assignments: subdomains 1, 2, 3, 4 correspond to HSC 70 domains IA, IB IIA, IIB. As these two structures are so similar we only mention the differences. Residues 350–373 in subdomain 1 of actin have no corresponding residues in the HSC 70 fragment. Subdomain 2 which binds to DNase I corresponds to the conserved (black) part in Fig. 3a of Flaherty *et al.* The unconserved (white) part of IB is replaced in actin by a loop containing the methylated histidine 73. In subdomain 4, a helix (Phe 223–Ala 230) is missing in the corresponding domain IIB. In addition,

the loop (residues 274–282) is absent in domain II. Finally, there seems to be a difference in the conformation of the phosphates and the divalent metal as found in HSC 70 and actin. HSC 70 is thought to be involved in ATP-dependent disassembly of clathrin cages as well as in transmembrane targeting of proteins to cellular organelles. Representatives of the family of the heat shock proteins have been found in *Escherichia coli* which are about 48% homologous to the eukaryotic proteins. The striking similarity of the three-dimensional structures of the ATPase fragment of HSC 70 and actin suggests the existence of a common prokaryotic ancestor. □

Received 18 May; accepted 25 July 1990.

1. Straub, F. B. *Studies, University of Szeged II*, 3–15 (1942).
2. Straub, F. B. *Studies, University of Szeged III*, 23–27 (1943).
3. Pollard, T. D. *Curr. Opin. Cell Biol.* **2**, 33–40 (1990).
4. Vandekerckhove, J. *Curr. Opin. Cell Biol.* **2**, 41–50 (1990).
5. Elzinga, M., Collins, J. H., Kuehl, W. M. & Adelstein, R. S. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2687–2691 (1973).
6. Collins, J. H. & Elzinga, M. *J. biol. Chem.* **250**, 5915–5920 (1975).
7. Lu, R. C. & Elzinga, M. *Biochemistry* **16**, 5801–5806 (1977).
8. Vandekerckhove, J. & Weber, K. *Eur. J. Biochem.* **90**, 451–462 (1978).
9. Lazarides, E. & Lindberg, U. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4742–4746 (1974).
10. Mannherz, H. G., Barrington-Leigh, J., Leberman, R. & Pfaff, H. *FEBS Lett.* **60**, 34–38 (1975).
11. Hitchcock, S. E. *J. biol. Chem.* **255**, 5668–5673 (1980).
12. Mannherz, H. G., Kabsch, W. & Leberman, R. *FEBS Lett.* **73**, 141–143 (1977).
13. Sugino, H. *et al. J. Biochem.* **86**, 257–260 (1979).
14. Suck, D., Kabsch, W. & Mannherz, H. G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4319–4323 (1981).
15. Kabsch, W., Mannherz, H. G. & Suck, D. *EMBO J.* **4**, 2113–2118 (1985).
16. Suck, D., Oefner, C. & Kabsch, W. *EMBO J.* **3**, 2423–2430 (1984).
17. Oefner, C. & Suck, D. *J. molec. Biol.* **192**, 605–632 (1986).
18. Kabsch, W. *J. appl. Crystallogr.* **21**, 916–924 (1988).
19. Kabsch, W. *J. appl. Crystallogr.* **21**, 67–71 (1988).
20. Kabsch, W., Pai, E. F., Mannherz, H. G. & Suck, D. *Springer Series in Biophysics* (eds Aebi U. & Engel, J.) **3**, 42–47 (1989).
21. Bruenger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
22. Read, R. J. *Acta crystallogr.* **A42**, 140–149 (1986).
23. Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1978).
24. Steitz, T. A., Fletterick, R. J., Anderson, W. A. & Anderson, C. M. *J. molec. Biol.* **104**, 197–222 (1976).
25. Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
26. Naharro, G., Robbins, K. C. & Reddy, E. P. *Science* **223**, 63–66 (1984).
27. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. *Nature* **347**, 44–49 (1990).
28. Polzar, B., Nowak, E., Goody, R. S. & Mannherz, H. G. *Eur. J. Biochem.* **182**, 267–275 (1989).
29. Suck, D., Lahm, A. & Oefner, C. *Nature* **332**, 464–468 (1988).
30. Boyer, M., Rouston, C. & Benyamin, Y. *Biosci. Rpt* **5**, 39–46 (1985).
31. Sutoh, K. *Biochemistry* **23**, 1942–1946 (1984).
32. Mornet, D. & Ue, K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3680–3684 (1984).
33. Mornet, D., Bonet, A., Audemard, E. & Bonicel, J. *J. Muscle Res. Cell Motil.* **10**, 10–24 (1989).
34. Jacobson, G. R. & Rosenbusch, J. P. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2742–2746 (1976).
35. Johnson, P., Wester, P. J. & Hikida, R. S. *Biochim. biophys. Acta* **578**, 253–257 (1979).
36. Konno, K. *J. Biochem., Tokyo* **103**, 386–392 (1988).
37. Schwyter, D., Phillips, M. & Reiser, E. *Biochemistry* **28**, 5889–5895 (1989).
38. Carlsson, L. *et al. J. molec. Biol.* **105**, 353–366 (1976).
39. Schütt, C. E., Lindberg, U., Myslík, J. & Strauss, N. *J. molec. Biol.* **209**, 735–746 (1989).
40. Vandekerckhove, J. S., Kaiser, D. A. & Pollard, T. D. *J. Cell Biol.* **109**, 619–626 (1989).
41. Kuwayama, H. & Yount, R. G. *Biophys. J.* **49**, 454a (1986).
42. Hegyi, Szilagy, L. & Elzinga, M. *Biochemistry* **25**, 5793–5798 (1986).
43. Saenger, W. *Principles of Nucleic Acid Structure* 74 (Springer-Verlag, 1984).
44. DePhamphilis, M. L. & Cleland, W. W. *Biochemistry* **12**, 3714–3723 (1973).
45. Strzelecka-Golaszewska, H., Boguta, G., Zmorzynski, S. & Moraczewska, J. *Eur. J. Biochem.* **182**, 299–305 (1989).
46. Nowak, E., Strzelecka-Golaszewska, H. & Goody, R. S. *Biochemistry* **27**, 1785–1792 (1988).
47. Loscalzo, J. & Reed, G. H. *Biochemistry* **15**, 5407–5412 (1976).
48. Brauer, M. & Sykes, B. D. *Biochemistry* **21**, 5934–5939 (1982).
49. Miki, M. & Wahl, P. *Biochim. biophys. Acta* **828**, 188–195 (1985).
50. Valentin-Ranc, C. & Carlier, M.-F. *J. biol. Chem.* **264**, 20871–20880 (1989).
51. Moore, P. B., Huxley, H. E. & DeRosier, D. J. *J. molec. Biol.* **50**, 279–295 (1970).
52. Sutoh, K. *Biochemistry* **21**, 3654–3661 (1982).
53. Bertrand, R. *et al. Biochemistry* **27**, 5728–5736 (1988).
54. Méjean, C. *et al. Biosci. Rpt* **6**, 493–499 (1986).
55. Miller, L., Kalnowski, M., Yunossi, Z., Buisinski, J. C. & Reiser, E. *Biochemistry* **26**, 6064–6070 (1987).
56. Moir, A. J. G., Levine, B. A., Goodearl, A. J. & Trayer, I. P. *J. Muscle Res. Cell Motil.* **8**, 68–69 (1988).
57. Barden, J. A., Phillips, L., Cornell, B. A. & dos Remedios, C. G. *Biochemistry* **28**, 5895–5901 (1989).
58. Barden, J. A. & Phillips, L. *Biochemistry* **29**, 1348–1354 (1990).
59. Szilagy, L. & Lu, R. C. *Biochim. Biophys. Acta* **709**, 204–211 (1982).
60. Trayer, I. P., Trayer, H. R. & Levine, B. A. *Eur. J. Biochem.* **164**, 259–266 (1987).
61. Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. *Nature* **346**, 623–628 (1990).
62. Blum, M., Metcalf, P. & Wiley, D. C. *J. appl. Crystallogr.* **20**, 235–247 (1987).
63. dos Remedios, C., Miki, M. & Barden, J. A. *J. Muscle Res. Cell Motil.* **8**, 97–117 (1987).
64. Priestley, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).

SUPPLEMENTARY INFORMATION. Requests should be addressed to the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. We thank G. A. Petsko for help with the refinement and R. S. Goody for discussions and comments on this manuscript. The support of the Deutsche Forschungsgemeinschaft is acknowledged.

Atomic model of the actin filament

Kenneth C. Holmes, David Popp, Werner Gebhard & Wolfgang Kabsch

Max-Planck-Institut für Medizinische Forschung, Abteilung Biophysik, Jahnstraße 29, 6900 Heidelberg, FRG

The F-actin filament has been constructed from the atomic structure of the actin monomer to fit the observed X-ray fibre diagram from oriented gels of F-actin. A unique orientation of the monomer with respect to the actin helix has been found. The main interactions are along the two-start helix with a contribution from a loop extending across the filament axis provided by the molecule in the adjacent strand. There are also contacts along the left-handed genetic helix.

ACTIN filaments (F-actin) are important components of the cell cytoskeleton where they are often involved in transport processes. In muscle tissue of all kinds it is the cyclical interaction between myosin, often also organized in filaments, and the thin filament, of which the major component is F-actin, which leads to contraction. Straub¹ originally isolated the actin monomer (G-actin) from muscle and showed that raising the salt con-

centration leads to the formation of F-actin polymers. The monomer binds one molecule of ATP which is hydrolysed to ADP on polymerization. Elongation of the actin filament proceeds at both ends but at different rates. The polymerization process has been studied in detail (see refs 2, 3). F-actin can be complexed with myosin subfragment 1 (S1) resulting in arrow-head-shaped filaments⁴. It has been shown that the pointed and barbed ends of the S1-decorated filaments correspond to the slow and fast polymerizing ends of F-actin, respectively⁵.

The structure of F-actin has been determined by electron microscopy (see ref. 6). It is helical with 13 actin molecules per 6 (left-handed) turns and a repeat of about 360 Å. As the rotation per molecule is 166°, the actin helix morphologically appears as two right-handed steep helices which twine slowly round each other.

The relevance and integrity of F-actin as a model of the thin filament is shown by its ability to stimulate the myosin ATPase^{7,8} and to enable myosin transport at high velocity in model systems⁹.

The structure of monomeric actin complexed with DNase I is reported in the accompanying paper¹⁰. It consists of two

domains which are known as large and small, although their difference of size is now known to be not great, the smaller containing both N and C termini. Each of the domains is further subdivided into two. ATP or ADP and a calcium ion are bound between the two domains.

F-actin can be oriented in capillary tubes¹¹ and the resulting oriented gel gives detailed X-ray fibre diffraction patterns which diffract to better than 10 Å resolution (Fig. 1). A number of important biological structures, notably DNA, have been determined from X-ray fibre diagrams by building a model and checking the model by calculating its diffraction pattern and comparing it with the observed fibre diffraction pattern¹². An optimization is achieved by trial and error or, when the model is good enough, by iterative least squares. We have used this approach in the present case to determine the structure of F-actin from the structure of monomeric actin and the fibre diffraction pattern of F-actin. Our working assumption is that the structures of F-actin and of actin as found in the complex with DNase I are the same. The model of the actin filament thus depends on only four free parameters: orientation and distance of the actin monomer with respect to the helix coordinate frame. Our assumption is born out by the goodness of fit we achieve.

The search

The atomic coordinates (ADP form) of the actin monomer reported in the accompanying paper¹⁰ were completed by adding the missing three C-terminal residues built as an extension of the C-terminal α -helix. Then the actin monomer was placed in the F-actin helix in all possible orientations in 10° steps. The calculated X-ray fibre diffraction intensities¹³ were compared with the intensities derived for layer-lines 0–21 out to a radial resolution of 16.6 Å. To this radial resolution it is possible to calculate the layer-line intensities by deconvolution¹⁴. For each

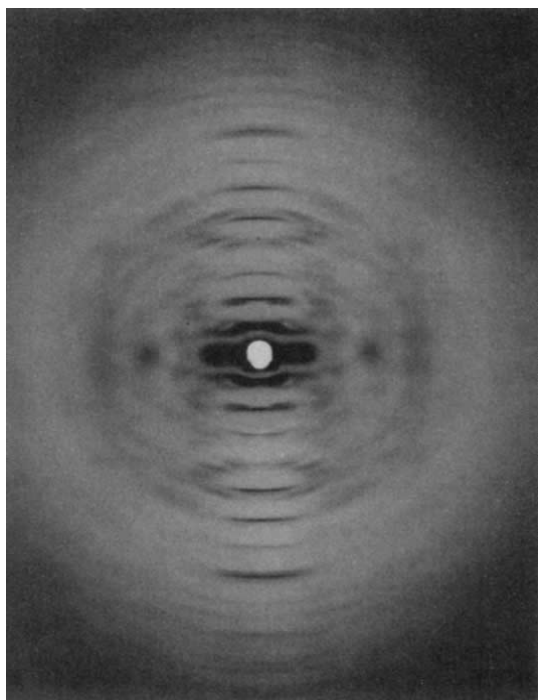


FIG. 1 X-ray fibre diagram from an oriented F-actin gel¹¹ obtained with CuK α radiation from a fine focus rotating anode tube (Elliott-GEC, GX18, 40 kV, 40 mA, 100 $\times$ 1,000 μ m focus) and a double focusing mirror (2 $\times$ 20 cm) optical system. Exposure time was 48 h. Specimen to film distance was $\sim$ 20 cm. A four-film pack (CEA REFLEX 25) was used to cover the wide dynamic range of the data. The layer-lines index on a 358 Å repeat with 13 subunits in 6 left-handed turns. Examination of the low order equatorial diffraction shows that the arrangement of the actin fibres is essentially a two-dimensional liquid with a minimum particle separation of about 130 Å. The F-actin used for this photograph was stabilized with phalloidin.

orientation, the radius of gyration^{15,16} was adjusted to 25 Å and a weighted residual, mean value 54% (Table 1), was calculated. We also computed the crystallographic modulus R-factor but found the residual to be more discriminating. The best solutions were common to both searches.

One solution gave a significantly better fit (residual 27%) than all other solutions. Only two other solutions gave residuals low enough to need to be considered (30% and 32%). Of these two solutions, the next best is an orientation close to the best solution which was found in earlier work using the low resolution electron density map from the crystal structure determination¹⁷. The structure of the third best solution is roughly a mirror image of the best solution. In all cases the large domain is near the helix axis.

Gels were made from actin complexed with phalloidin (1:1) to stabilize the F-actin filament. Gels of actin without phalloidin give X-ray fibre diagrams which are difficult to distinguish from gels with phalloidin except that the presence of phalloidin makes longer exposures possible. Phalloidin was not, however, included in the model calculations. To try to estimate what effect the omission of phalloidin (MW 789) might have, we omitted 10 different heptapeptides from the model and recomputed the residual of the best structure: changes of 1–3% were obtained. These are small enough not to have invalidated our search strategy.

The next best solution, which is about 20° from our best solution was favoured in earlier work¹⁷ using the 6-Å resolution electron density map. This solution, however, would not refine satisfactorily against high angle data. The reason for its then being favoured over the present best solution lies in the critical role of solvent contrast and contrast variation, which was not taken account of in the earlier work.

To discriminate between the three best solutions they were refined against high resolution data by the method of least squares. At resolutions better than 16 Å the closely spaced layer-lines and disorientation (s.d. 2.8° for our best sample) make deconvolution for the layer-line intensities a hazardous procedure. Therefore, at high resolution we have not attempted to deconvolute the layer-lines but rather have calculated the disoriented diffraction pattern. A comparison of the observed and calculated intensities over the whole film has been used as the basis of a least squares refinement for the orientation and radial position of the actin monomer (4 d.f.).

Using data out to 8.4 Å resolution a best fit was found (Fig. 2) which was derived from the best solution given by the low angle search. This was reproducibly reached from starting positions up to 10° away from the final orientation. It was also stable with respect to radial displacements of 1 Å or more. The final radius of gyration was 23.5 Å. The overall residual was 15.6% (R-factor 22%). None of the other solutions from the low angle search refined to a residual better than 27.0% against high resolution data, demonstrating that the best solution is highly significant.

Moreover, for the best solution there is no collision of main chain atoms with the main chain of neighbouring molecules in the helix. Collisions of side chain atoms can be relieved by reorientation of side chains. For the next best orientation, overlap of main chain atoms occurs in the region of the 10-residue loop (264–273) which protrudes from the surface (see ref. 10). For the third best orientation there is extensive interpenetration of neighbouring molecules. Therefore, the best fit is favoured by an independent stereochemical criterion.

The overall fit to the X-ray fibre diagram is very good: the intensities of the strong low order layer-lines are well accounted for as are some notable absences such as the J_6 -term on the third layer-line. At high resolution, peak intensities are not all accurately predicted but in practically every case an observed peak is matched by a calculated peak. The changes of intensity of some of the high order peaks may be attributable to the presence of phalloidin which has not been included in the calculation. In spite of the fact that the actin monomer has been

taken as a rigid body and that at this stage we have made no attempt to let the domains move or to improve the model in any other way, a good fit to the observed diffraction pattern can be achieved using only four parameters.

Domain movement

Some changes of structure between monomeric and F-actin are to be expected however. Moreover, the loss of phosphate from ATP is a recognizable step in the maturation of the actin helix¹⁸ which may be accompanied by structural changes. The question is how big these changes might be.

As the initial fit to the diffraction pattern is rather good we can try and estimate the magnitude of the changes to the starting model necessary to fit the X-ray pattern optimally by investigating the effects of domain movement on the diffraction pattern. Four domains have been defined as the subdomains 1–4 in ref. 10. Comparison with the domain motions in hexokinase shows the domain boundaries to be plausible. Using the method outlined above, we let the four domains refine separately (22 d.f. using the best fit as a starting position). The residual (for the whole pattern) dropped from 15.6% to 9.8% and the R-factor from 22% to 16.9%. The result was stable. Moreover, the domain movements were small (2–3°) except for subdomain 2, which twisted through about 15°. The largest shift was 6 Å. In all cases helical contacts were strengthened. In particular, the contact between subdomain 2 and the neighbouring large domain along the two-start helix became stronger.

The results strongly suggest that the adjustments to the initial best fit model necessary to fit the actin diffraction pattern are small. The larger relative motion of subdomain 2 could reflect its orientation being influenced by the binding to DNase I in the crystal structure.

Contrast variation

If one calculates the fibre diffraction pattern from the atomic coordinates without taking account of the water surrounding the fibre one is unable to scale the low resolution part of the pattern and the high resolution part of the pattern at the same time. This effect arises from a variation of contrast with resolution: at very low resolution the water may be treated as a continuum, which lowers the effective contrast; at high resolution there is no effect. In between, where we calculate, things are difficult to quantify. A solution to the problem was first proposed by Langridge *et al.*¹² for their work on DNA fibres. Each atom is considered to be surrounded by a hole into which water cannot penetrate. The effective scattering factor of the atom is then its normal electronic scattering factor less the scattering from a sphere full of water with the appropriate radius. We have used this approach: the 'hole' radii were taken to be the van der Waals radii for the given groups (methyl, methylene, etc.). We took the values and type definitions from the program QUANTA (Polygen Inc.). A factor was applied to the van der Waals radii to produce the observed average rise of intensity with resolution. A value of 0.98 was found by trial and error.

Structure of F-actin

The structure of the filament derived from the rigid body transformation of the ADP crystal coordinates is shown in Fig. 3. The differences produced by independent domain refinement do not affect our conclusions at the present level of description. The maximum diameter is between 90–95 Å. The large domain (comprising subdomains 3 and 4 in ref. 10) is at small radius and the small domain (subdomains 1 and 2) is at large radius. The large domain is about 55 Å in its long dimension and fits rather naturally along the long two-start actin helix. In the following discussion one should bear in mind that we have completed the C terminus (373–375) by model building. The interactions between molecules along the two-start actin helix are extensive, involving residues 322–325 with 243–245; 286–289 with 202–204; 166–169 and 375 with 41–45 (Fig. 3b). Some of

these interactions are mimicked by the DNase I-actin interactions. To this extent DNase I may be thought of as an actin analogue. The DNase I binding loop (41–50), which is hydrophobic, makes contact with the large domain of the molecule above. In the middle of the fibre the density is low. Contacts along the left-handed genetic helix are provided between residues 110–112 and 195–197. In addition, a loop from the opposite strand (Phe 266, Ile 267, Gly 268, Met 269, see below) appears to insert into the hydrophobic pocket formed by residues Tyr 166, Ala 169, Leu 171, His 173, Cys 285, Ile 289, Gly 63, Ile 64, and residues 40–45.

The loop (264–273) may readily be rebuilt as an antiparallel β sheet with a β bend in the middle. This generates a finger-like structure extending across the helix axis from the middle of the long α -helix. This rearrangement puts the sequence Phe-Ile-Gly-Met in the β bend in such an orientation that the hydrophobic residues can intercalate between the molecules of the opposing strand to form a dense hydrophobic core or 'plug'. On the basis of this rebuild one can propose a plausible model for the interactions between the molecules in F-actin: the strong interaction is brought about by the formation of an extensive hydrophobic core between neighbouring molecules along the two-start helix into which the 'plug' residues 269–272 from the molecule in the opposite strand are inserted. The interaction is three-bodied^{3,19,20}. If the plug were to be withdrawn it is easy to imagine that the remaining longitudinal interaction might not be strong enough to hold the helix together.

The hydrophobic interactions are the easiest to recognize. Nevertheless, some possible salt bridges can be identified: Arg 39 lies between Asp 286 and Glu 270 all on different molecules; Arg 205 on one molecule lies between Asp 244 and Glu 205 on the neighbouring molecule; Arg 62 is near to Asp 288 on its neighbour. Specific hydrogen bonds are to be expected but cannot be specified at present.

The N terminus, thought to be involved in myosin binding, is at the largest radius (~45 Å). Known myosin-binding residues (Asp 1, Glu 2, Asp 3, Glu 4, Asp 24, Asp 26, Arg 28, Glu 93, Arg 95, Lys 336) (see discussion in ref. 10) form a patch near the outside of the filament (Fig. 3).

Comparison with previous data

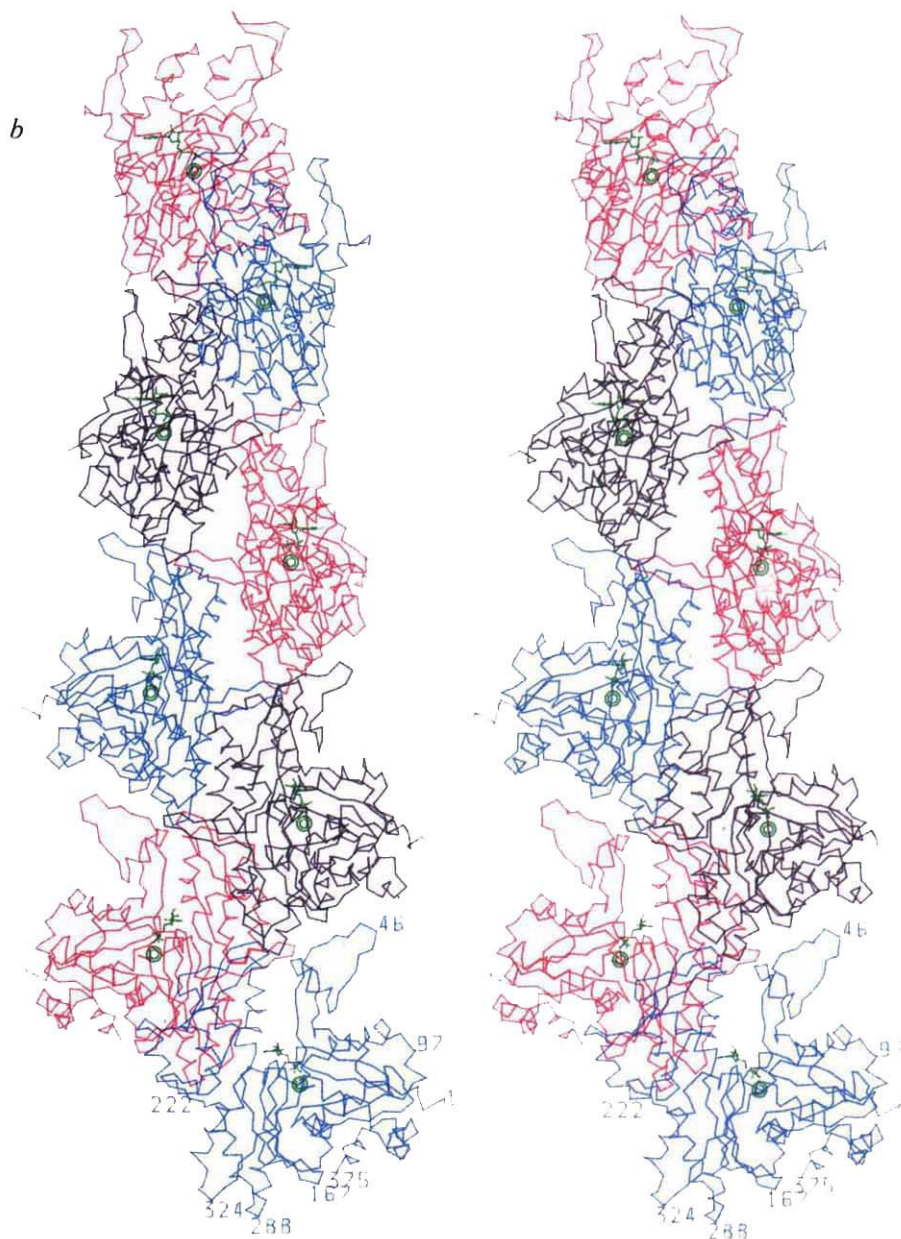
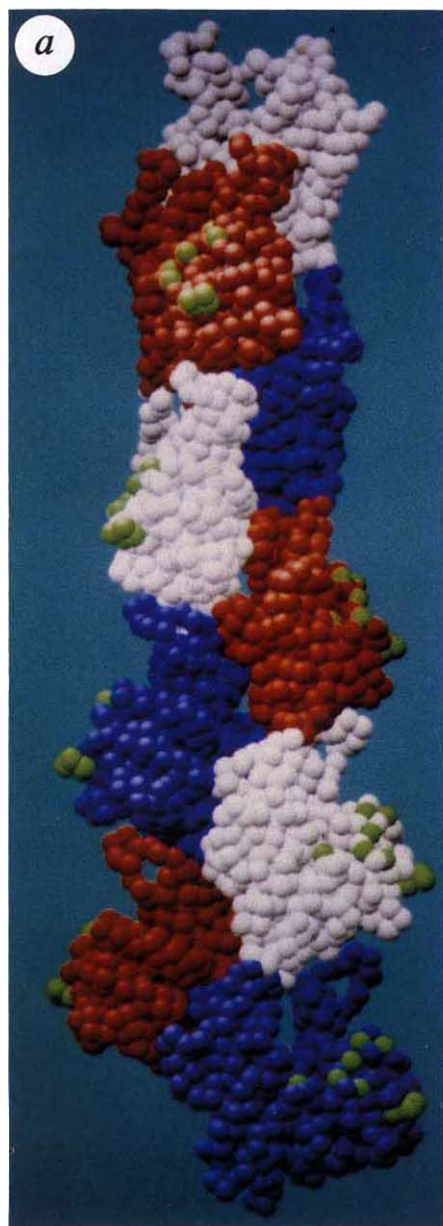
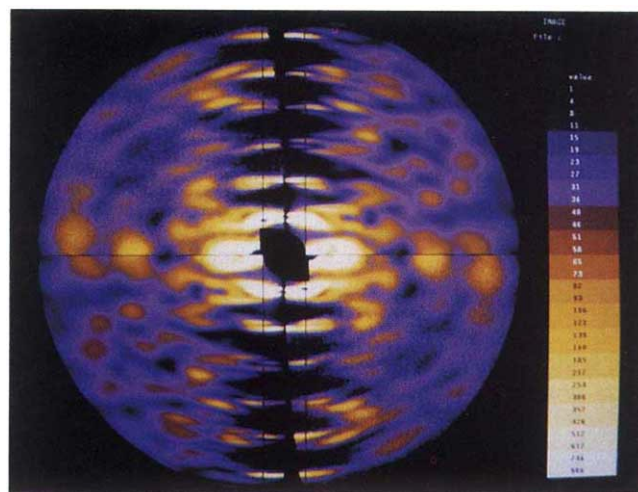
Our model of the filament shares many features with the electron microscopic image reconstructions of Mg-paracrystals of F-actin²¹ which show an elongated actin monomer with the long

FIG. 3 a Representation of eight monomers of the F-actin helix as a space filling model. For clarity each actin monomer is shown in a different colour. Each amino-acid residue is represented by a sphere of radius 2.7 Å. The small domain is at large radius. Amino acid residues that cross-link to myosin in the actomyosin complex are shown in green. The matrix and vector translation necessary to generate the molecule at the bottom of this figure from the orthorhombic crystal coordinates (ADP form) as deposited with the Brookhaven Protein Data Bank are:

$$\begin{pmatrix} -0.71034 & 0.16333 & 0.68465 \\ -0.58180 & -0.68369 & -0.44053 \\ 0.39614 & -0.71126 & 0.58068 \\ -3.77 & 131.21 & -37.97 \end{pmatrix}$$

whereby the helix axis is along Z. b, A stereo pair of the C $_{\alpha}$ -positions in the corresponding eight monomers. ADP and calcium are shown in green. A few amino-acid sequence numbers are given for the actin monomer at the bottom of the picture. This monomer is at the 'barbed' end of the filament. Strong contacts are made along the steep two-start helix. There are also contacts along the left-handed genetic helix. Note the loop (residues 262–272) crossing the helix axis. The hydrophobic residues Phe 266, Ile 267 (Gly 268) and Met 269 seem to insert into a hydrophobic cavity formed between neighbouring molecules on the opposite strand. In this way the stabilization of the longitudinal contact between two subunits along the two-start helix may be increased by the cooperation of a third subunit from the opposite strand.

FIG. 2 A comparison of the measured data (upper left, lower right) with the calculated diffraction pattern (upper right, lower left). Data were obtained from the films shown in Fig. 1 by densitometry (Optronics), scaling the films⁴⁵ and subtracting background as a polynomial, $a_1/s^4 + a_2/s^2 + a_3 + a_4s^2$, (where s is the radial distance in reciprocal space) fitted to points in between layer-lines. The calculated data was obtained by deriving the layer-line intensities¹³ for each order of Bessel function given by the selection rule $l = -6n + 13m$ using the full atomic coordinates of the model depicted with C α -positions in Fig. 3 and taking account of contrast variation as is described in the text. The intensities were convoluted with the gaussian function $f(Z) = \exp(-Z^2/2t_0^2)$ where $t_0 = 0.00071 \text{ \AA}^{-1}$ to take account of instrumental layer-line broadening. The broadening masks the effects of cumulative angular disorder⁴⁶ which would not affect our calculated patterns for values up to about 10–15°. A second convolution was made for disorientation, which was treated as a gaussian⁴⁷ with a disorientation parameter of 2.8°. The best fit of the model to the observed data was obtained by iterative least squares based on a point for point comparison of the observed and calculated diffraction patterns and the minimization of the residual defined in Table 1. Data out to the edge of the figure at 8.4 Å resolution gave an R -factor of 15.6%. The near meridional data out to 100 Å resolution, shown with two vertical lines, was excluded from the comparison because the gaussian approximation is not valid in this region.



axis roughly at right angles to the filament axis. This axis does not intersect the filament axis but passes at a distance of about 15 Å so as to leave a low density region in the centre of the filament. The maximum diameter of our filament is between 90–95 Å which is close to the values found by a number of authors (see ref. 6) and involves the residues of the N terminus. In agreement with Aebi *et al.*²² we find that the main contacts are along the long two-start helix. By comparison with S1 decorated actin filaments^{4,23} and knowledge of the S1 binding region in our model (Fig. 3a) we can deduce the polarity of the filament: the barbed end is at the bottom. Inspection of the reconstruction of decorated actin obtained in ice²³ shows longitudinal contacts and contacts along the left-handed genetic helix much as in the present model but the longitudinal contact appears the weaker. This could be an effect of resolution as although subdomain 2 makes extensive contacts with subdomain 3 and accounts for half the longitudinal contact, it has little mass. The contact along the left-handed genetic helix involves fewer residues but involves the bulky subdomain 4 being close to subdomain 3 of the neighbouring molecule. At limited resolution this juxtaposition would give the impression of a stronger contact than that mediated by the less substantial subdomain 2.

The assigned polarity of the model of the actin filament is in agreement with a number of biochemical observations. ADP-ribosylation of actin at Arg 177 by clostridial toxins blocks its polymerization^{24,25}. Arg 177 lies near the axis between the long two-start helix strands. A bulky group cannot be accommodated between the strands in this position. From Fig. 3b it is clear that the only position where the ADP-ribosylated actin could be accommodated is at the barbed end, namely at the bottom of the filament in the figure. This explains why ADP-ribosylated actin acts like a barbed end capping protein. In agreement with our model it has been shown²⁶ that profilin binds at the barbed end near the C terminus of actin at residue Glu 364. The actin molecule at the top in Fig. 3 at the pointed end of our filament model should be capable of binding DNase I. Indeed, Podolski and Steck²⁷ found that DNase I associates with the pointed ends of actin filaments in human red blood cell membrane skeletons.

Elzinga and Phelan²⁸ crosslinked residues Lys 191 and Cys 374 residing on different monomers in the filament. They concluded that the prosthetic groups of these residues are not more than 14 Å apart. Our value is 19 Å for the distance between the (model built) C_α 374 and C_α 191 on neighbouring molecules along the left-handed genetic helix. This agrees well with the observation.

Modification of His 40 prevents fibre formation²⁹, as does fluorescein labelling of Lys 61³⁰. Both these residues are involved in the extensive contacts between neighbouring molecules along the long two-start helix. Labelling of Tyr 53 stops polymerization³¹: this residue is not itself directly part of the actin-actin interface but appears to have an important role in stabilizing the 40–50 loop, which lies immediately adjacent to it. The labelling of Tyr 69 impairs but does not stop polymerization³²: the residue is close to an interface but is not directly involved so that the effects must be secondary. The reactivities of lysine residues 50, 61, 68, 113, 284 and 291 decrease on actin polymerization^{33,34}. Again, all these amino acids are located near contact regions between monomers in our model of the filament. Lys 113 is adjacent to the contacts along the left-handed genetic helix (110–112 with 195–197). It is conceivable that phalloidin which is known to bind in this region stabilizes this contact³⁵. In addition, it has been demonstrated that the mutated β actin (Gly 245 to Asp) has reduced ability to polymerize^{36,37}. The mutated residue is involved in interactions along the two-start actin helix in our model.

Yanagida and Oosawa³⁸ measured the fluorescence anisotropy of etheno-ADP bound in place of ADP in F-actin fibres. They found the chromophore dipoles to be at 75° to the axis. We measure the plane of the nucleotide to be at 86° to the

helix axis, which is compatible. The distance of the N6-atom of ADP from the filament axis is 21 Å in our model. This is in agreement with the observed value of 25 Å obtained from fluorescence energy transfer measurements³⁹.

Cys 274 is a favourite target for labelling with fluorescent dyes and for energy transfer measurements⁴⁰ in F-actin fibres. Thus Taylor *et al.*⁴¹ measured the radius of Cys 374 to be 37 Å and Kasprzak *et al.*⁴² obtain 20–25 Å. Our data actually stop with Arg 372, but if we extrapolate by continuing the α helix as has been described above the radius of the Cys 374 C_α is 25 Å. The radial coordinate suggested⁴² for Gln 41 (40 Å) is rather larger than we observe (Gln 41 O_γ, 26 Å). This discrepancy could arise from the length of the linker.

TABLE 1 Residual search of the actin monomer orientation

	Latitude									
	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
0°	46	45	45	52	53	46	53	52	52	52
10°	50	44	43	51	53	53	53	53	53	
20°	52	44	48	53	54	54	54	54	46	
30°	53	47	43	54	54	49	54	51	43	
40°	54	47	49	55	55	55	53	46	41	
50°	54	38	51	54	52	51	51	42	45	
60°	51	38	50	46	51	51	53	36		
70°	54	44	49	48	50	52	50	38		
80°	53	52	49	50	49	54	46	36		
90°	52	50	49	51	55	53	36	43		
100°	51	49	49	53	54	47	30	43		
110°	51	49	49	52	50	43	30	44		
120°	48	47	47	52	44	40	35			
130°	45	46	52	49	41	33	41			
140°	45	48	52	44	38	31	45			
150°	50	51	50	43	36	27	42			
160°	51	49	47	38	35	32	42			
170°	48	45	42	35	38	41	49			
180°	44	39	37	33	43	43				
190°	39	34	34	33	47	49				
200°	37	33	32	36	46	51				
210°	40	35	32	41	42	52				
220°	43	38	36	46	50	48				
230°	41	44	42	49	51					
240°	38	44	47	45	52					
250°	40	44	49	47	53					
260°	47	48	51	48	52					
270°	53	52	51	49						
280°	50	51	50	48						
290°	48	48	49	50						
300°	48	47	48	52						
310°	50	46	47							
320°	50	44	47							
330°	50	48								
340°	50	50								
350°	48									

Shown are the percentage values of the residual $R = \sum_i w_i (l_{\text{calc}} - l_{\text{obs}})^2 / \sum_i w_i l_{\text{obs}}^2$, where $w_i = 1/(l_{\text{obs}} + 40)$ (40 is the average value of the background) calculated for 22 layer-lines out to 16 Å resolution for all orientations of the ADP-actin monomer in 10° steps. l_{obs} was obtained by deconvolution by the method of Makowski¹⁴ from the data shown in Fig. 2. l_{calc} was calculated from the atomic coordinates for each orientation of the actin monomer using the method of Klug *et al.*¹³ taking account of contrast variation as described in the text. For each orientation the radius of the actin monomer from the helix axis was adjusted to make the radius of gyration 25 Å. The latitude and longitude of a chosen axis through the centre of gravity of the monomer are shown as ordinate and abscissa. The molecule was spun around this axis by an angle ω in 10° intervals and the smallest value of R obtained is shown in the table. The three best solutions are: 27% (lat.=50°, long.=150°, ω =100°); 30% (lat.=60°, long.=110°, ω =80°); 32% (lat.=20°, long.=210°, ω =30°). The first two solutions remained distinct on refinement at high resolution. Solutions near the third refine into a single solution.

Discussion

There is one orientation (and only one) of the actin monomer as derived from the actin-DNase I complex which accounts for the fibre diffraction pattern. In fact it accounts for the diffraction pattern rather well. To obtain a residual of 15.6% with only four degrees of freedom is surprisingly good. The solution provides a fairly accurate road map for the F-actin structure. It is not our intention at this stage to describe all the stabilizing interactions in detail, although some have already become apparent.

Furthermore, the orientation of the actin monomer is consistent with a number of independent biochemical and biophysical measurements which had no role in the determination. *Post facto*, the cited measurements are sufficient to determine the

orientation of the actin monomer without recourse to the X-ray diffraction pattern, although it is not clear that one would have been able to accomplish this synthesis without having an independent criterion for selection of facts, as not all published results form an internally consistent set.

The differences between the structures derived from actin complexed with DNase I to monomeric actin and actin in F-actin seem to be small. This could imply that the DNase I complex is an analogue of F-actin because the DNase I binding mimics one set of actin-actin interactions. Alternatively, the structures of F- and G-actin may be very similar. We expect that a comparison with the structurally closely related heat shock protein HSC 70⁴³ and with the actin-profilin complex⁴⁴ may shed light on this matter. □

Received 18 May; accepted 25 July 1990.

1. Straub, F. B. *Studies, University of Szeged II*, 3-15 (1942).
2. Pollard, T. D. *Curr. Opin. Cell Biol.* **2**, 33-40 (1990).
3. Korn, E. D. *Physiol. Rev.* **62**, 672-737 (1982).
4. Moore, P. B., Huxley, H. E. & DeRosier, D. J. *J. molec. Biol.* **50**, 279-295 (1970).
5. Woodrum, D. T., Rich, S. A. & Pollard, T. D. *J. Cell Biol.* **67**, 231-237 (1975).
6. Egelman, E. H. *J. Muscle Res. & Cell Motil.* **6**, 129-151 (1985).
7. Taylor, E. W. *Crit. Rev. Biochem.* **6**, 103-164 (1979).
8. Adelstein, R. S. & Eisenberg, E. A. *Rev. Biochem.* **49**, 921-956 (1980).
9. Spudich, J. A., Kron, S. J. & Sheetz, M. P. *Nature* **315**, 584-586 (1985).
10. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. *Nature* **347**, 37-44 (1990).
11. Popp, D., Lednev, V. V. & Jahn, W. *J. molec. Biol.* **197**, 679-684 (1987).
12. Langridge, R. et al. *J. molec. Biol.* **2**, 38-64 (1960).
13. Klug, A., Crick, F. H. C. & Wyckoff, H. W. *Acta crystallogr.* **11**, 199-213 (1958).
14. Makowski, L. *J. appl. Crystallogr.* **11**, 273-283 (1978).
15. Hartt, J. & Mendelson, R. *Fedn Proc.* **39**, 1728 (1980).
16. Curmi, P. M., Stone, D. B., Schneider, D. K., Spudich, J. A. & Mendelson, R. A. *J. molec. Biol.* **203**, 781-798 (1988).
17. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. *Springer Series in Biophysics* (eds Aebi, U. & Engel, J.) **3**, 48-50 (1989).
18. Carlier, M.-F. *Int. Rev. Cytol.* **115**, 139-170 (1989).
19. Oosawa, F. & Kasai, M. *J. molec. Biol.* **4**, 10-21 (1962).
20. Erickson, H. P. *J. molec. Biol.* **206**, 465-470 (1989).
21. O'Brien, E. J., Couch, J., Johnson, G. R. P. & Morris, E. P. In *Actin* (eds dos Remedios, C. G. & Barden, J. A.) 3-15 (Academic, New York, 1983).
22. Aebi, U., Millonig, R., Salvo, H. & Engel, A. *Ann. N.Y. Acad. Sci.* **483**, 100-119 (1986).
23. Milligan, R. A. & Flicker, P. F. *J. Cell Biol.* **105**, 29-39 (1987).
24. Aktories, K. & Wegner, A. *J. Cell Biol.* **109**, 1385-1387 (1989).

25. Vandekerckhove, J., Schering, B., Bärmann, M. & Aktories, K. *FEBS Lett.* **225**, 48-52 (1987).
26. Vandekerckhove, J. S., Kaiser, D. A. & Pollard, T. D. *J. Cell Biol.* **109**, 619-626 (1989).
27. Podolski, J. L. & Steck, T. L. *J. biol. Chem.* **263**, 638-645 (1988).
28. Elzinga, M. & Phelan, J. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6599-6602 (1984).
29. Hegyi, G., Premecz, G., Sain, B. & Muhrad, A. *Eur. J. Biochem.* **44**, 7-12 (1974).
30. Burtnick, L. D. *Biochim. biophys. Acta* **791**, 57-62 (1984).
31. Bender, N., Fasold, H., Kenmoku, A., Middelhoff, G. & Volk, K. E. *Eur. J. Biochem.* **64**, 215-218 (1976).
32. Lehrer, S. S. & Elzinga, M. *Fed. Proc.* **31**, 502 (1972).
33. Lu, R. C. & Szilagyi, L. *Biochemistry* **20**, 5914-5919 (1981).
34. Hitchcock-De Gregori, S. E., Mandal, S. & Sachs, G. A. *J. biol. Chem.* **257**, 12573-12580 (1982).
35. Vandekerckhove, J., Deboen, A., Nassal, M. & Wieland, T. *EMBO J.* **4**, 2815-2818 (1985).
36. Vandekerckhove, J., Leavitt, J., Kakunaga, T. & Weber, K. *Cell* **22**, 893-899 (1980).
37. Taniguchi, S., Sagara, J. & Kakunaga, T. *J. Biochem., Tokyo* **103**, 707-713 (1988).
38. Yanagida, T. & Oosawa, F. *J. molec. Biol.* **126**, 507-524 (1978).
39. Miki, M., Hamby, B. D. & dos Remedios, C. G. *Biochim. biophys. Acta* **871**, 137-141 (1986).
40. dos Remedios, C., Miki, M. & Barden, J. A. *J. Muscle Res. Cell Motil.* **8**, 97-117 (1987).
41. Taylor, D., Reidler, L., Spudich, J. A. & Stryer, L. *J. Cell Biol.* **89**, 362-367 (1981).
42. Kasprzak, A. A., Takahashi, R. & Morales, M. F. *Biochemistry* **27**, 4512-4522 (1988).
43. Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. *Nature* **346**, 623-628 (1990).
44. Schutt, C. E., Lindberg, U., Myslik, J. & Strauss, N. *J. molec. Biol.* **209**, 735-746 (1989).
45. Fox, G. C. & Holmes, K. C. *Acta crystallogr.* **20**, 886-891 (1966).
46. Egelman, E. H. & DeRosier, D. J. *Acta crystallogr.* **A38**, 796-799 (1982).
47. Holmes, K. C. & Barrington Leigh, J. *Acta crystallogr.* **A30**, 635-638 (1974).

ACKNOWLEDGEMENTS. We thank R. Mendelson for many helpful discussions. We also thank E. O'Brien and members of his team at Kings College London for collaboration in an earlier stage of this work.

The identity and position of X-ray sources in globular clusters: radio emission from NGC6712

Harry J. Lehto, Graham Machin, Ian M. McHardy & Paul Callanan

Department of Astrophysics, Keble Road, University of Oxford, Oxford OX1 3RH, UK

NINE globular clusters are known to contain luminous ($L_X > 10^{36} \text{ erg s}^{-1}$) X-ray sources, but because the X-ray position error boxes, a few arcseconds in size, contain many stars, optical identification of the sources has not been possible, with the exception of that in NGC7078 (ref. 19). This has prevented detailed investigation of the emitting objects. If the X-ray sources are similar to low-mass X-ray binaries (LMXBs) outside globular clusters, low-level radio emission would be expected^{1,2}, and the detection of such radio emission would both strengthen the case that luminous globular cluster X-ray sources are LMXBs and aid optical identification by means of an accurate radio position. We report here the detection of radio emission from the position of the luminous X-ray source A1850-08 (ref. 3) in the globular cluster NGC6712. The centroid of the slightly elongated radio source coincides with a previously suggested 20th magnitude stellar optical counterpart⁴.

If the radio emission is synchrotron radiation, the minimum energy requirement and magnetic field strength are similar to what is found in the radio counterparts to other high-luminosity X-ray sources such as Sco X-1 and Cyg X-3.

As well as NGC6712, we have observed three other globular clusters⁵ to search for radio emission from the X-ray sources. The globular cluster NGC6712 is at a distance of 6.5 kpc (ref. 4). Observations³ made in 1974 with Ariel-5 revealed an X-ray source, A1850-08, associated with the cluster. This source is also listed in the ANS⁶, SAS-3⁷ and Uhuru⁸ surveys (4U1850-08). A1850-08 is an X-ray burster^{8,9} and has an average quiescent X-ray flux density of $4 \mu\text{Jy}$ at 5.2 keV (refs 10, 11). Detailed observations by Grindlay *et al.*¹² with the Einstein high-resolution imager provide the best available X-ray position ($18^{\text{h}} 50^{\text{m}} 21.18^{\text{s}}$, $-8^{\circ} 46' 04.4''$ with a 3σ accuracy of $\sim 2.5''$). Recently Cudworth⁴ obtained photographic plates of the region with subarcsecond resolution and suggested as a possible counterpart a $V=20$ magnitude ($40 \mu\text{Jy}$) blue star at $18^{\text{h}} 50^{\text{m}} 21.13^{\text{s}}$, $-8^{\circ} 46' 04.2''$, 0.8 arcsec from the X-ray position. Bailyn *et al.*¹³ reported independently the detection of a faint ($B > 20$) very blue object (star X) located 0.6 arcsec southwest of Cudworth's star (C. D. Bailyn, personal communication). Very recently Nieto *et al.*¹⁴ have reported high-resolution sub-pupil photon-counting imaging of the X-ray error box, and have detected two $B=21$ magnitude stars (which they call S and Z), about 0.6-0.7 arcsec southwest of Cudworth's star. The exact positions of these stars will be discussed in future publications

by Nieto *et al.* and Bailyn *et al.* Deep U-band CCD imaging of the area, also by Nieto *et al.*¹⁴, shows the existence of a very blue object which may be either star S or Z, and may coincide with star X (C. D. Bailyn, personal communication).

We observed NGC6712 for 84 min in February 1989 at 6 cm with the Very Large Array (VLA) in a hybrid A/B configuration. The continuum bandwidth of 100 MHz was used. Other technical details are given in ref. 5. The data were edited and calibrated at the VLA with the late DEC-10 computer. The maps were made with AIPS on the Oxford University Astrophysics Department VAX3800 computer. CLEANing and using a taper at 200k resulted in a r.m.s. noise level of $\sigma = 24 \mu\text{Jy}$ per beam, which was calculated from 43,776 source-free pixels. A 51.2×51.2 arcsec map (pixel size = $0.2''$) centred on the X-ray position was made. The only source above a $\sim 3.5\sigma$ level is a 6.7σ source centred at $18^{\text{h}} 50^{\text{m}} 21.08^{\text{s}}$, $-8^{\circ} 46' 04.18''$ (1950) with a peak flux density of $163 \mu\text{Jy}$ per beam and an integrated flux density of $195 \pm 23 \mu\text{Jy}$ (Fig. 1). The CLEANed components were convolved with a circular beam with a half-power beam width (HPBW) of 2.00 arcsec, which is equal to the mean of the major and minor axes of the dirty beam.

The position of this new radio source is consistent with the Einstein position¹², and its luminosity is similar to that of the radio sources associated with LMXBs outside clusters. As the chance coincidence of such a radio source with the X-ray error box is small¹⁵, we are confident that the radio and X-ray source are physically related and therefore that the more accurate radio position can be used to search for the optical counterpart to the X-ray source.

The proposed optical counterparts (stars S, Z and X and Cudworth's star) all lie within 0.8 arcsec of the radio centroid. The formal positional error of the fit to the radio position is < 0.1 arcsec but we estimate that the error for absolute radio astrometry is about 0.4 arcsec. The uncertainties in the optical positions are also about 0.4 arcsec, and so all the candidate stars are in agreement with our position. Based on their colours, the very blue object of Nieto *et al.*¹⁴ and star X (C. D. Bailyn, personal communication) are the most promising candidates at

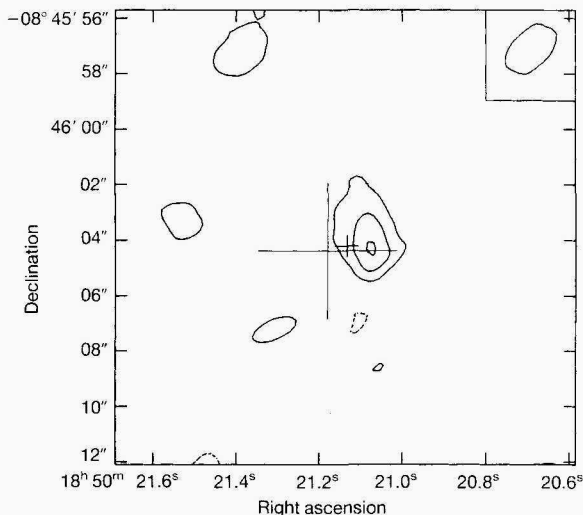


FIG. 1 The 6-cm VLA map of NGC6712. The central 16.4×16.4 arcsec are shown. The contours are $(-1, 1, 2, 3)$ times $52.5 \mu\text{Jy}$ per beam. The r.m.s. noise level of the map is $24.3 \mu\text{Jy}$ per beam. The peak flux density is $163 \mu\text{Jy}$ per beam (6.7σ) and the integrated flux density is $195 \mu\text{Jy}$. The half-power contour of the dirty beam is shown in the box. Note that the major axis of the dirty beam is in a direction almost perpendicular to the observed source extension, confirming that the extension of the source is real. A circular restoring beam with a HPBW of 2 arcsec was used to make the map. The X-ray position is shown, as is that of Cudworth's star⁴. The optical identifications by Bailyn *et al.*¹³ and Nieto *et al.*¹⁴ are not shown as their coordinates have not been published, but they are stated to lie 0.6–0.7 arcsec southwest of Cudworth's star, well within the radio contours.

present, but final identification must await more detailed optical spectroscopy or photometry.

The radio brightness distribution is well represented by a point source with one-sided extended emission. A $2.1''$ extension in position angle -11° , unresolved ($\leq 0.3''$) in its perpendicular direction (79°) and convolved with the beam reproduces the structure of this extended emission. Assuming a characteristic width of 0.3 arcsec, a length of 2 arcsec, and the most likely random angle (30°) with the line of sight, we estimate the volume of the emitting region to be $V \approx 3.3 \times 10^{50} \text{ cm}^3$. The emitting gas fills a volume $k_V V$, where $k_V \leq 1$ is a scaling factor. As probable emission mechanisms from this volume, we consider bremsstrahlung and synchrotron.

If the radiation is bremsstrahlung, the temperature of the radiating plasma has to be less than $\sim 6 \times 10^5 \text{ K}$; otherwise we would expect to see an extended optical emission region of integrated magnitude 19, which would have been detected in the deep images of Cudworth⁴. Heating by the central source¹⁶ sets a lower limit for the temperature at about 10^4 K . The observed radio flux density ($195 \mu\text{Jy}$) requires a particle density of $n = 5.1 \times 10^2 k_V^{-1/2} T^{1/4} \geq 5 \times 10^3 \text{ cm}^{-3}$. Here we have made the standard assumption that the mean square of the plasma atomic number, $\langle Z^2 \rangle$, is 1.4 and have used a Gaunt factor $\bar{g} = 1.2$. If we also require the cooling time of the plasma to be at least the light crossing time of the system ($6.7 \times 10^6 \text{ s}$), we obtain an upper limit for the density of $n \leq 1.3 \times 10^4 T^{1/2}$. The scaling factor has respective values of $1 \geq k_V \geq 2 \times 10^{-6}$. The total mass of the emitting cloud is at least $M \geq 1.4 \times 10^{-6} M_\odot$.

If the mechanism were synchrotron radiation, with a spectral shape of $S \propto \nu^{-0.8}$ between 10 MHz and 100 GHz, then the flux density from the observed volume implies¹⁷ an equipartition magnetic field of $B_{\text{min}} = 1.7 k_V^{2/7} \text{ mG}$, assuming the ratio of proton to electron energy density is $\epsilon_p/\epsilon_e = 100$. The minimum total energy required is $E_{\text{min}} = 8.8 \times 10^{43} k_V^{3/7} \text{ erg}$ and the lifetime of the source¹⁷ is $\tau = 1.0 \times 10^{12} \text{ s} = 3.2 \times 10^4 \text{ yr}$, which is much larger than the light crossing time of the source. Thus there is no need to invoke particle reacceleration in the source.

We note that synchrotron-powered double radio lobes have been detected around the high-luminosity X-ray sources Sco X-1, Cyg X-3 and SS433. The linear size we see here (2 light months) is a factor of ten smaller than those seen in the above sources, but the magnetic field strength that we derive here is similar to that found in those sources¹⁸. Also the minimum energy required here is close to that required in SS433, although in Cyg X-3 it is two orders of magnitude larger, and in Sco X-1 it is smaller by a factor of 30. One distinct possibility is therefore that the radio source associated with A1850–08 is a smaller version of the synchrotron-powered double radio sources found around some other high-luminosity X-ray sources.

Further observations to determine the radio spectral index would enable us to distinguish between the two possible emission mechanisms, and higher-resolution observations are required to determine whether we are observing a point source with a jet, a multiple point-source structure, or a point source together with truly extended emission. $\square$

Received 13 February; accepted 20 July 1990.

- Hjellming, R. M. & Johnston, K. J. in *The Physics of Accretion onto Compact Objects* (eds Mason, K. O., Watson, M. G. & White, N. E.) (Springer, New York, 1986).
- Grindlay, J. E. & Seaquist, E. R. *Astrophys. J.* **310**, 172–175 (1986).
- Seward, F. D., Page, C. G., Turner, M. J. L. & Pounds, K. A. *Mon. Not. R. astr. Soc.* **175**, 39P–46P (1976).
- Cudworth, K. M. *Astr. J.* **96**, 105–115 (1988).
- Machin, G., Lehto, H. J., McHardy, I. M., Callanan, P. J. & Charles, P. A. *Mon. Not. R. astr. Soc.* (in the press).
- Grindlay, J. E. *et al. Astrophys. J.* **212**, L67–L71 (1977).
- Doxey, R., Apparo, K. M. V., Brandt, H. V., Dower, R. G. & Jernigan, J. G. *Nature* **269**, 112–116 (1977).
- Cominsky, L., Forman, W., Jones, C. & Tananbaum, H. *Astrophys. J.* **211**, L9–L13 (1977).
- Swank, J. H., Becker, R. H., Pravdo, S. H., Saba, J. R. & Serlemitsos, P. J. *IAU Circ. No.* 3010 (1976).
- Jernigan, J. F. & Clark, G. W. *Astrophys. J.* **231**, L125–L129 (1979).
- Brandt, H. V., Doxsey, R. E. & Jernigan, J. G. in *Proc. Symp. 21st Plenary Mtg of COSPAR* (eds Baity, W. A. & Peterson, L. E.) 3–66 (Pergamon, New York, 1978).
- Grindlay, J. E., Hertz, P., Steiner, J. E., Murray, S. S. & Lightman, A. P. *Astrophys. J.* **282**, L13–L16 (1984).

13. Baily, C. D., Grindlay, J. E., Cohn, H. & Lugger, P. M. *Astrophys. J.* **331**, 303–312 (1988).
14. Nieto, J. L. *et al.* *IAU Circ. No.* 4944 (1990).
15. Fomalont, E. B., Kellerman, K. I., Wall, J. V. & Weistrop, D. *Science* **225**, 23–28 (1984).
16. Kalman, T. R. & McCray, R. *Astrophys. J. Suppl. Ser.* **50**, 263–318 (1982).
17. Pacholczyk, A. G. *Radio Astrophysics* (Freeman, San Francisco, 1970).
18. Strom, R. G., van Paradijs, J. & van der Klis, M. *Nature* **337**, 234–236 (1989).
19. Aurère, M., Le Fèvre, O. & Terzan, A. *Astr. Astrophys.* **138**, 415–420 (1984).

Regular magnetic fields in coronae of spiral galaxies

Dmitry Sokoloff* & Anvar Shukurov†

* Physics Department, Moscow State University, Moscow 119899, USSR
† IZMIRAN, Academy of Sciences, Troitsk 142092, USSR

THE idea that large-scale ordered magnetic fields might reside in the gaseous coronae of spiral galaxies came first from theoretical arguments¹ that such fields could be generated *in situ* by turbulent hydromagnetic dynamos. Here we put this general argument into practice. The partially ionized gas in galactic coronae is in chaotic motion, both turbulent and convective, and owing to galactic rotation acquires a non-zero mean helicity. Differential rotation and this helicity generate a magnetic field with predicted magnitude $\sim 1 \mu\text{G}$ and with a scale comparable to that of the corona. This field is nearly independent of the disk magnetic field: the two have different symmetry with respect to the galactic equator, the disk field being even and the corona field odd. Consequently, a galactic-scale neutral sheet arises $\sim 1 \text{ kpc}$ above one side of the disk. At heights of about 3–5 kpc, the poloidal magnetic field component dominates the corona, a result consistent with observations of polarization in the galaxy NGC4631.

The presence of magnetic fields in the coronae of at least some galaxies is firmly established. These fields manifest themselves through synchrotron emission originating far above the plane of the Milky Way^{2–5} and external galaxies seen edge on (ref. 6). Considerable contribution to the synchrotron emission probably comes from chaotic magnetic fields the scale of which, 0.3–1 kpc, is much below the corona size (the typical radius of the coronae, R , is 10–15 kpc). The origin of this chaotic magnetic field can be associated both with field advection from the disk by galactic-fountain motions or by Parker instability (typical scale of 1 kpc) and with *in situ* generation (Section VII.13 in ref. 1). Recently, polarization observations of edge-on galaxy NGC4631 (ref. 7) have revealed the presence of a regular magnetic field at the height of several kiloparsecs above the galactic plane.

The dynamo mechanism, which can be efficient in coronae of spiral galaxies, is the mean-field turbulent dynamo discussed in detail in refs 1, 8–10. A dynamo of this type requires the presence of helical chaotic motions. Convective chaotic motions are present in the coronae owing to galactic fountains¹¹ or Parker instability⁸. Hydrodynamic turbulence, which is probably maintained in coronae¹², also contributes to the mean helicity. Chaotic motions in any rotating and density-stratified fluid body become helical, that is, they are characterized by a non-vanishing correlator $\mathcal{H} = \langle \mathbf{v} \cdot \nabla \times \mathbf{v} \rangle$ called the mean helicity, where angular brackets denote averaging and $\mathbf{v}$ is the chaotic velocity field. In physical terms, this means that vorticity in turbulent cells ($\nabla \times \mathbf{v}$) has a non-vanishing average projection onto the direction of their motion, $\mathbf{v}$, that is, trajectories of fluid elements become helical and the total number of right-hand helices differs from the number of left-hand ones. The mean helicity arises because of the action of the Coriolis force: the turbulent cells that move outwards, where the density ρ is smaller, expand and thereby acquire an extra component of angular velocity $\nabla \times \mathbf{v}$ which in the northern hemisphere is antiparallel to the angular velocity of overall rotation $\boldsymbol{\omega}$. The cells that fall toward the centre, however, are compressed and the associated extra vorticity is

parallel to $\boldsymbol{\omega}$. Thus, the scalar product $\mathbf{v} \cdot \nabla \times \mathbf{v}$ has some dominant sign for all turbulent cells in a given hemisphere and its mean value is non-zero. The resulting mean helicity is proportional to $\boldsymbol{\omega} \cdot \nabla \rho$.

A non-vanishing mean helicity is a typical property of cosmic bodies that rotate and are stratified. Correspondingly, the mean-field turbulent dynamos are common in astrophysical environments. There is no doubt that the mean helicity is non-zero in both the coronae and disks of spiral galaxies but there have been no attempts to estimate it from, for example, observations of chaotic velocities of interstellar gas in the Milky Way.

The generation of a regular magnetic field in a turbulent helical flow of conducting fluid is governed by the averaged induction equation^{1,8–10}

$$\frac{\partial \mathbf{B}}{\partial t} = R_\omega \nabla \times (\mathbf{V} \times \mathbf{B}) + R_\alpha \nabla \times (\alpha \mathbf{B}) + \nabla^2 \mathbf{B} \quad (1)$$

where $\mathbf{B}(\mathbf{r}, t)$ is the mean magnetic field, $\mathbf{V}(\mathbf{r})$ is the regular velocity field, represented in coronae primarily by the overall differential rotation, $\mathbf{V}(\mathbf{r}) = \boldsymbol{\omega}(\mathbf{r}) \times \mathbf{r}$, and $\alpha(\mathbf{r})$ is the helicity coefficient proportional to $\mathcal{H}$. Equation (1) is written in the dimensionless form with distances and time measured in units of R and of magnetic diffusion time R^2/β , respectively, where β is the turbulent magnetic diffusivity. The functions of position $\alpha(\mathbf{r})$ and $\boldsymbol{\omega}(\mathbf{r})$ are normalized by their extremal values α_0 and ω_0 which are absorbed into dimensionless numbers R_α and R_ω . These latter numbers are in fact turbulent magnetic Reynolds numbers, which measure the intensity of the dynamo sources, the mean helicity and differential rotation

$$R_\alpha = \frac{\alpha_0 R}{\beta} \quad \text{and} \quad R_\omega = \frac{\omega_0 R^2}{\beta} \quad (2)$$

Usually the dynamos are characterized by two related quantities, R_α^2 and the dynamo number $D \equiv R_\alpha R_\omega$. For a given R_α^2 , the growing or steady-state non-trivial solutions to equation (1) appear for sufficiently large dynamo numbers, $D > D_{\text{cr}}$.

Now we estimate R_α and the critical dynamo number D_{cr} for galactic coronae and compare them with the expected values of D . At moderate heights above galactic disks ($\lesssim 5 \text{ kpc}$) the regular velocity field preserves cylindrical symmetry and only weakly depends on the height z . Variation of V with z can only facilitate generation of magnetic fields. It seems plausible that the rotation curve is flat at large distances from the rotation axis in the corona as well as in the disk. Thus, the rotation curve may be chosen as $V(\mathbf{r}) = V_0[1 - \exp(-s/s_0)]$, where s is the cylindrical radius and V_0 and s_0 are numerical constants. Correspondingly, $\omega_0 = V_0/s_0$. We take the rotation velocity in the corona to be of the same order of magnitude as in the disk, typically $V_0 \approx 200 \text{ km s}^{-1}$. It is well known^{1,8,9} that $\alpha(\mathbf{r})$ is positive in the northern hemisphere and negative in the southern one. We choose the simplest form $\alpha(\mathbf{r}) = \cos \theta$, where θ is the spherical polar angle. Then the approximate solution of equation (1) using the Galerkin expansion in the free-decaying modes¹³ shows that, for $R_\alpha = 3$, the critical dynamo number D_{cr} is ~ 200 for $s_0/R = 0.1$, 300 for $s_0/R = 0.3$, and 1,500 for $s_0/R = 1$. These numbers are probably underestimates and are accurate within a factor of three or five.

We now estimate D for galactic coronae. The maximal value of the helicity coefficient can be estimated as $\alpha_0 \approx l^2 \omega_0 h^{-1}$ ($\propto \boldsymbol{\omega} \cdot \nabla \rho$), where l is the turbulent scale and h is the density scale height (section VI.4 in ref. 1). As follows from the mixing-length theory, $\beta \approx lv/3$. Thus

$$R_\alpha \approx 3 \frac{V_0}{v} \frac{Rl}{s_0 h} \quad R_\omega \approx 3 \frac{V_0}{v} \frac{R^2}{ls_0}$$

$$D \approx 9 \left(\frac{V_0}{v} \right)^2 \frac{R^3}{s_0^2 h}$$

As a plausible estimate of the scale of chaotic motions we adopt $l \approx 0.5 \text{ kpc}$. For the r.m.s. turbulent velocity $v = 100 \text{ km s}^{-1}$ and

$h = 3$ kpc (the values typical of the Milky Way corona) we obtain $R_\alpha \approx 3$ and $D \approx 2,000$ for $s_0/R = 0.3$ and $R = 15$ kpc.

The order-of-magnitude estimates given in the previous paragraphs imply that in galactic coronae the dynamo number is generally close to or exceeds D_{cr} . Parameters of real galaxies have a natural scatter around the typical values and we conclude that self-excitation of regular coronal magnetic fields may not occur in all galaxies.

Although the adopted value $R = 15$ kpc might seem to be too large, this choice is motivated by the fact that this is the radius of the supernovae distribution in the disk of the Milky Way and it is the supernovae that provide the hot gas that fills the corona. For $R = 10$ kpc we obtain $D \approx 600$, which is also sufficient for the dynamo action.

As examples, we estimate D and D_{cr} for NGC891 and NGC4631, the two edge-on galaxies for which the radiohaloes have been investigated observationally. For illustration purposes, we use the parameters of rotation curves in disks of these galaxies. For NGC891, $V_0 \approx 246$ km s $^{-1}$ and $s_0 \approx 8.9$ kpc (see ref. 14). Thus, $R_\alpha \approx 2$, $D \approx 770$ and $D_{cr} \approx 800$, and $D \approx D_{cr}$ within the accuracy of our estimates. Similarly, for NGC4631 we have

$V_0 \approx 150$ km s $^{-1}$ and $s_0 \approx 3.2$ kpc which gives $R_\alpha \approx 3$, $D \approx 2,200$ and $D_{cr} \approx 300$. We see that the dynamo can act in the coronae of both galaxies, although it is weaker in NGC891.

We emphasize that in galactic coronae, as for many other quasi-spherical dynamo objects, the dominant magnetic mode is axisymmetric and its φ - and r -components have odd symmetry with respect to the equator: $B_{r,\varphi}(r, \theta) = -B_{r,\varphi}(r, \pi - \theta)$. In the galactic disks, however, the regular magnetic field has even symmetry^{1,8,10}. This allows us to predict a galactic-scale neutral sheet extending parallel to the galactic plane. On the one side of the disk, the regular fields of the disk and the corona are directed similarly and give rise to a relatively strong field. On the other side of the disk these two fields have opposite directions and a neutral sheet arises. The sheet appears between the fields of scales $h \approx 3$ kpc and $z_0 \approx 0.5$ kpc (the ionized disk half-thickness). The sheet thickness should be comparable to the smaller of these scales and the energy release is as low as $\approx 2.5 \times 10^{11}$ erg s $^{-1}$. Nevertheless, the neutral sheet can be observed through asymmetry of the synchrotron intensity and of the Faraday rotation measure above and below the galactic disk surrounded by the coronal magnetic field. The distributions of Faraday rotation measures of extragalactic radio sources are notably different in the northern and southern galactic hemispheres¹⁵; this can be interpreted as an indication of the neutral sheet in the northern galactic hemisphere. An asymmetry in the distribution of synchrotron intensity above and below the disk has been observed in the corona of NGC891 (U. Klein, personal communication).

Observations of the corona of NGC4631 indicate that the regular magnetic field has a dominant poloidal component although at some positions toroidal and poloidal components are comparable⁷. It is also well known that turbulent dynamos generate predominantly toroidal magnetic fields for $R_\omega \gg R_\alpha$, which is true for galactic coronae. For instance, in the model of ref. 13 the energy stored in poloidal field is only 1/9 of that in the toroidal one. For odd symmetry, however, the toroidal field vanishes at the equator and the poloidal field dominates for $75^\circ \leq \theta \leq 105^\circ$, or for $z \leq 3$ –5 kpc (Fig. 1b).

The steady-state strength of the regular magnetic field can be estimated^{1,10} from the balance of the Lorentz force produced by the field and the Coriolis force which makes the chaotic motions helical

$$\frac{B_r B_\varphi}{4\pi h} \approx \rho v \omega_0 \quad (3)$$

For $B_r/B_\varphi \approx 1/3$ and gas density $\rho \approx 1.7 \times 10^{-27}$ g cm $^{-3}$, we obtain $B \approx 2$ μ G. The actual field strength can be below this value by a factor of three. For example, a similar relation yields $B \approx 7$ μ G in the Milky Way disk whereas the observed value is ~ 2 μ G. Thus, $B \approx 0.5$ –1 μ G is a plausible estimate of the regular field strength in galactic coronae. The magnetic energy density corresponding to $B = 1$ μ G, 4×10^{-14} erg cm $^{-3}$ is comparable to the densities of both thermal and turbulent kinetic energies, $nkT \approx \rho v^2/2 \approx 10^{-13}$ erg cm $^{-3}$. The predicted regular magnetic field should therefore be an important dynamic component of the corona. In particular, it can affect chaotic motions thereby reducing the mean helicity to such a level that the growing field becomes marginally stable and equation (3) holds. The role of the regular magnetic field in the corona dynamics deserves detailed study.

For $D \approx D_{cr}$, the steady-state magnetic field distribution is close to the marginally stable eigenfunction¹⁶, which can be approximated by a linear combination of free-decaying modes¹³. The toroidal field strength reaches the maximum for polar angles $\theta \approx 45^\circ$ and $\theta \approx 135^\circ$. The distributions of toroidal and poloidal magnetic fields in the marginally stable eigenfunction for $R_\alpha = 3$ and $s_0 = 0.3$ are shown in Fig. 1.

The synchrotron emissivity $\varepsilon \propto n_e H^2$ of the Milky Way corona is 10% of the disk emissivity² (here n_e is the relativistic electron density). If the total magnetic field H is twice the regular field

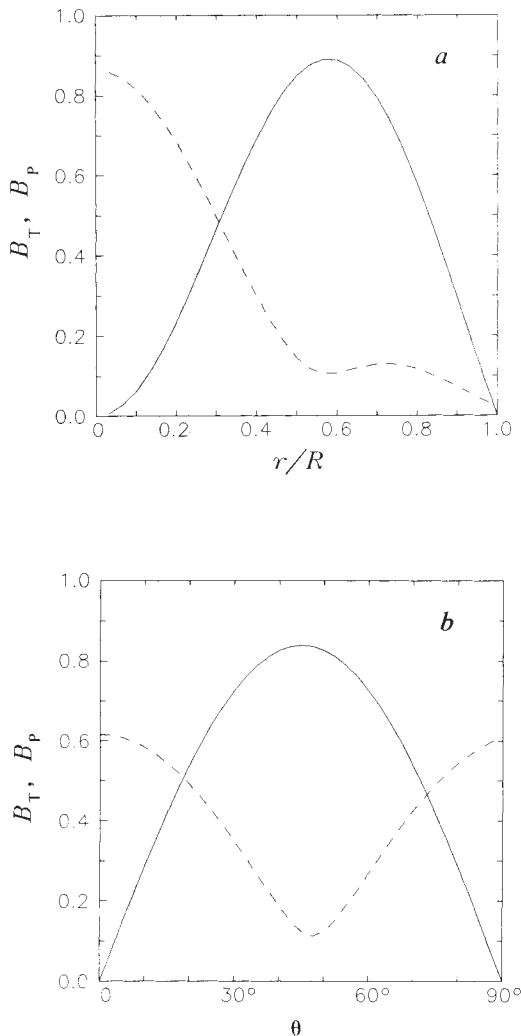


FIG. 1 Normalized strengths of toroidal (B_T , solid lines) and poloidal (B_P , dashed lines) components of the steady-state magnetic field as functions of *a*, the spherical radius at the polar angle $\theta = 45^\circ$ and *b*, θ for $r/R = 2/3$. The typical field strength is ~ 0.5 –1 μ G. The plots are based on approximate solution of the coronal dynamo problem¹³ and illustrate the general features of the spatial distributions rather than the detailed behaviour. For the field in the southern hemisphere $B_T(180^\circ - \theta) = -B_T(\theta)$. The field distributions are distorted by the disk field at $r \cos \theta \leq 1$ kpc which is not shown here.

($H \approx 4 \mu\text{G}$ in the disk and $H \approx 2 \mu\text{G}$ in the corona), the density of cosmic-ray electrons in the corona is smaller than the disk value by a factor of 2.5.

A regular coronal magnetic field of strength $1 \mu\text{G}$ can confine cosmic-ray protons up to $\approx 1.5 \times 10^{19}$ eV (that is, for a Larmor radius smaller than the corona radius). Even the space surrounding magnetized coronae is permeated by a dipole magnetic field, which decreases as r^{-3} . This vacuum field can further improve the confinement of cosmic rays. The problem of the origin of ultra-high-energy cosmic rays should be reconsidered taking into account the possible regular magnetic field in the galactic corona. □

Received 10 July; accepted 19 July 1990.

1. Ruzmaikin, A. A., Shukurov, A. M. & Sokoloff, D. D. *Magnetic Fields of Galaxies* (Kluwer, Dordrecht, 1988).
2. Badwar, G. D. & Stephens, S. A. *Astrophys. J.* **212**, 494–506 (1977).
3. Webster, A. *Mon. Not. R. astr. Soc.* **185**, 507–520 (1978).
4. Phillips, S., Kearsy, S., Osborne, J. L., Haslam, H. & Stoffel, H. *Astr. Astrophys.* **103**, 405–414 (1981).
5. Beuermann, K., Kanbach, G. & Berkhuijsen, E. M. *Astr. Astrophys.* **153**, 17–34 (1985).
6. Sofue, Y., Fujimoto, M. & Wielebinski, R. *A. Rev. Astr. Astrophys.* **24**, 459–498 (1986).
7. Hummel, E., Lesh, H., Wielebinski, R. & Schlickeiser, R. *Astr. Astrophys.* **197**, L29–L31 (1988).
8. Parker, E. N. *Cosmical Magnetic Fields* (Clarendon, Oxford, 1979).
9. Krause, F. & Rädler, K.-H. *Mean-Field Magnetohydrodynamics and Dynamo Theory* (Pergamon, Oxford, 1980).
10. Ruzmaikin, A., Sokoloff, D. & Shukurov, A. *Nature* **336**, 341–347 (1988).
11. Shapiro, P. R. & Field, G. B. *Astrophys. J.* **205**, 762–765 (1976).
12. Savage, B. D. & Massa, D. *Astrophys. J.* **314**, 380–396 (1987).
13. Sokoloff, D. & Shukurov, A. *Geophys. Astrophys. Fluid Dyn.* (submitted).
14. Kyazumov, G. A. *Astr. Zh. (SSSR)* **61**, 846–853 (1984).
15. Ruzmaikin, A. A., Sokoloff, D. D. & Kovalenko, A. V. *Soviet Astr.* **22**, 395–401 (1978).
16. Kvasz, L., Sokoloff, D. & Shukurov, A. *Geophys. Astrophys. Fluid Dyn.* (submitted).

Evaporation kinetics of forsterite and implications for the early solar nebula

Akihiko Hashimoto

Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA

THE planets and meteorites originated in a disk of gas and dust—the solar nebula—in which thermal processing fractionated the elements to varying degrees between gaseous and condensed phases^{1,2}. A record of these effects is preserved in the chemical, mineralogical and isotopic compositions of meteorites, which accreted in the nebula. Previous attempts to interpret these fractionation effects have been based largely on thermodynamic calculations^{3,4}, which assume that phases remained in equilibrium with one another. Vapour–solid equilibria in the system Mg–Si–O–H have also been studied experimentally⁵. But equilibrium is only a first approximation to the real situation: internal evidence in the chondritic meteorites⁶ shows that the thermal processing was rapid, and kinetic effects (non-equilibrium fractionation) must have played an important part in the evaporation and condensation events that created the chondrules, refractory inclusions and dust grains that are preserved in chondrites. Here I report the results of experiments to study the kinetics of evaporation of Mg_2SiO_4 (forsterite), the most abundant mineral in planets and meteorites. Solid and liquid Mg_2SiO_4 evaporate very slowly, only about one-tenth as fast as they would if the vapour species produced by evaporation were those predicted by equilibrium thermodynamics, and if no energy barrier in excess of the energy of reaction impeded the process. This large kinetic effect is due mainly to the transient production of gaseous SiO_2 on evaporation. As a result, forsterite may evaporate at the same rate as an intrinsically more refractory material.

The processes governing rates of evaporation and condensation depend on the nature of the material. In the case of metals, all vapour molecules that impinge on a solid or liquid surface

condense⁷. Thus the condensation rate of a metal vapour in equilibrium with its own condensed phase is equal to the rate of the incident molecules, J_e , which can be calculated from the kinetic theory of gases⁸

$$J_e = P(2\pi MRT)^{-1/2} \text{ mol cm}^{-2} \text{ s}^{-1} \quad (1)$$

where P is the equilibrium vapour pressure, M the molecular weight, R the gas constant, and T the temperature. At equilibrium the evaporation rate of atoms is equal to the condensation rate from a saturated vapour, by definition; therefore the evaporation rate of a metal is also J_e . But in the case of the silicates and oxides, the principal components of planets and meteorites, evaporation and condensation are much more complex processes because they involve the decomposition and/or recombination of various molecules and atoms⁹. In this circumstance evaporation and condensation rates are not simply equal to J_e . These processes need to be studied experimentally.

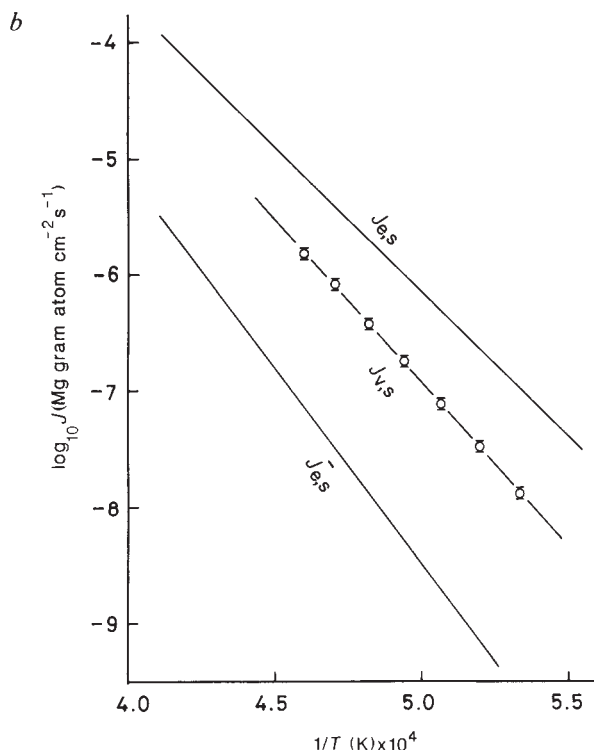
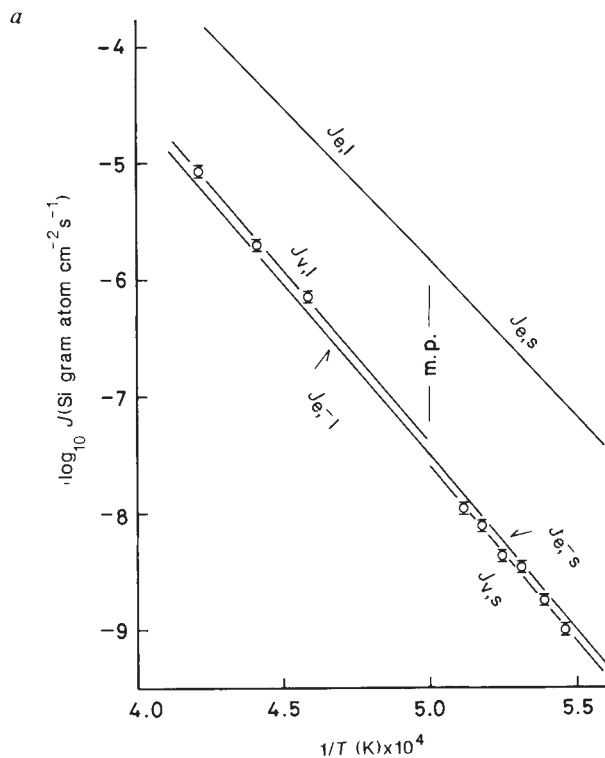
It is important to be able to separate evaporation from condensation so they can be studied individually. Evaporation without condensation occurs, and can be studied, in a vacuum. The evaporation rate of a particular material in a vacuum is called its free evaporation rate, J_v (ref. 8). Although J_v is measured in a vacuum, it is the same rate of evaporation that occurs at equilibrium; in the latter case it is simply compensated by the equal condensation rate. Because the condensation rate cannot exceed the incident rate of vapour molecules, J_v cannot exceed J_e ; and therefore J_e is regarded as the theoretical maximum evaporation rate. Evaporation is often described in terms of the ratio J_v/J_e . This evaporation coefficient, α_v , cannot exceed unity for the reason just given. In situations where kinetic factors are unimportant, α_v should tend to unity.

Evaporation experiments were carried out in a vacuum furnace (height, 18 inches; inside diameter, 14 inches) equipped with a turbomolecular pump and temperature control, which can operate up to $2,500^\circ\text{C}$ at 1×10^{-6} mm Hg. Each sample (100–200 mg) to be evaporated from the solid state was pressed into a disk (12 mm in diameter, ~ 1 mm thick) and sintered at $1,450^\circ\text{C}$ in air, then suspended by a tungsten–rhenium wire inside a pair of hemicylindrical tungsten heaters in the furnace. Samples to be evaporated from the molten state were pressed into cylinders (6 mm in diameter, ~ 4 mm high) and sintered, then suspended in iridium wire loops. A seven-layer tungsten–molybdenum thermal radiation shield surrounding the heaters kept the hot zone at a uniform temperature. Multiple holes in the radiation shield permitted efficient evacuation of the vapour generated from the sample; the vapour pressure in the vicinity of the sample was maintained at less than one-tenth of its intrinsic (or Langmuir, defined below) vapour pressure. Temperature was measured with a W/W26Re thermocouple and one-colour optical pyrometer; results agreed to $\pm 5^\circ\text{C}$. During each run the temperature was brought to the desired value quickly (20 – $100^\circ\text{C min}^{-1}$) and maintained for a prescribed time (3 – 500 min). The heating power was then switched off and the residual charge cooled quickly by radiation. After the run, the weight loss, surface area and chemical composition of the residual charge were determined, and the evaporation rate was calculated from these and the time at temperature.

Traditionally, pains are taken to control the oxygen fugacity of the environment in which experiments on mineral systems are carried out. The concepts of oxygen fugacity and redox potential are meaningful only for an equilibrium system, however, not for a non-equilibrium evaporation controlled by kinetics. The (dynamic) partial pressure of oxygen produced by evaporation of oxides, which may be defined in this case, is not an experimentally controllable factor, but an intrinsic property of the evaporation kinetics.

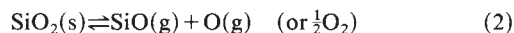
Because forsterite is compounded of MgO and SiO_2 , I first studied the evaporation of these simple oxides. Experiments showed α_v to be 0.12 – 0.22 for solid MgO (periclase) at $1,600$ – $1,900^\circ\text{C}$, consistent with previous work¹⁰. For SiO_2 , α_v is much

smaller, 0.011–0.015 for the solid (high cristobalite) at 1,560–1,685 °C and 0.038–0.048 for the liquid at 1,900–2,100 °C. (For solid SiO_2 , $\alpha_v \approx 0.01$ has been determined previously by the effusion-cell method^{11,12}, but with a large uncertainty.) For both MgO and SiO_2 , α_v increases with temperature. There is a linear relationship between $1/T$ and $\log \alpha_v$, because both $\log J_v$ and $\log J_e$ are linearly related to $1/T$ (Fig. 1).



The values of α_v cited above are dependent on values of J_e . The SiO , O_2 , O , SiO_2 , Si and Si_2 gases (in order of decreasing abundance) are in equilibrium with solid and liquid SiO_2 . For each of these species, the equilibrium incident rate (in $\text{mol cm}^{-2} \text{s}^{-1}$) can be calculated from equation (1). From these the total incident rate of Si-bearing molecules, expressed in terms of Si gram-atom $\text{cm}^{-2} \text{s}^{-1}$, can be calculated. The J_e curve in Fig. 1 was obtained in this way, and used with the J_v curve (from experimental measurements) to derive the small values of α_v given above for Si evaporation.

One can also determine α_v for each postulated evaporation reaction taken alone. For example,



As evaporation is a reaction that proceeds forward when expressed as above (the backward reaction is condensation), the fastest path determines the evaporation rate of $\text{SiO}_2(\text{s})$; this must be identical to the free evaporation rate referred to here, because J_v represents the rate of the forward reaction. If there is no kinetic energy barrier attached to this reaction path, J_v should be equal to the value of J_e calculated for the Si vapour species produced by this fastest reaction ($\alpha_v = 1$). If there is a kinetic barrier, on the other hand, J_v should be smaller than this J_e ($\alpha_v < 1$). $\text{SiO}(\text{g})$ is the predominant Si species in the vapour in equilibrium with solid or liquid SiO_2 , and it therefore controls the position of J_e in Fig. 1a. The fact that $\alpha_v \ll 1$ when calculated with this value of J_e implies either that the forward

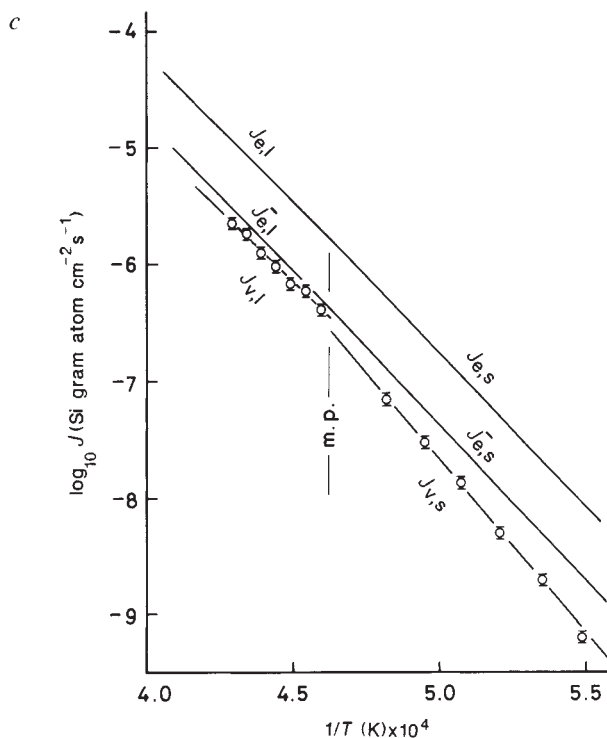


FIG. 1 Evaporation rates as a function of $1/T$, for a, SiO_2 ; b, MgO and c, Mg_2SiO_4 . $J_{v,s}$ and $J_{v,l}$ are the free evaporation rates from solid and liquid phases, respectively. $J_{e,s}$ and $J_{e,l}$ are the theoretical maximum condensation (or evaporation) rates of Si or Mg atoms, whether as monatomic or molecular species, onto (or from) solid and liquid surfaces of phases in equilibrium with the vapour. J_{e-s} and J_{e-1} are also such rates, but calculated on the basis of specific reaction mechanisms that involve gaseous SiO_2 and MgO . All molecules in the vapour phase are assumed to be generated only by evaporation from the condensed phase.

reaction (2) is kinetically hindered, or that reaction (2) is not in fact the fastest reaction path.

It is interesting to explore this second possibility. Curve J_e in Fig. 1a represents the incident rate for a gas of pure SiO_2 (instead of SiO) that is in equilibrium with solid and liquid SiO_2 . For these values of J_e , α_v becomes 0.7–0.9 for solid SiO_2 , almost unity. Similarly, α_v becomes 1.2 for molten SiO_2 . Thus there is no need to invoke a kinetic energy barrier, if reaction (3) is the actual rate-determining path for evaporation. (The fact that α_v exceeds unity for the case of evaporation from a liquid—which is impossible, according to the definition of α_v —may be due to underestimation of the surface area of the experimental charges.) By converting the experimental values for J_v through equation (1) into corresponding pressures (called Langmuir vapour pressure), the activation enthalpy and activation entropy can be calculated from the slope and intersection (with the ordinate), respectively, of the curve in a plot of $1/T$ against $\ln P$. The values obtained, $\Delta H^* = 138.0 \pm 13.7$ (2σ) kcal mol⁻¹ and $\Delta S^* = 38.3 \pm 7.2$ (2σ) cal mol⁻¹ K⁻¹, agree closely with the enthalpy of vaporization ($\Delta H^\circ = 139.7 \pm 8.0$) and entropy of vaporization ($\Delta S^\circ = 39.6 \pm 0.8$) calculated by equilibrium thermodynamics¹³ for the reaction $\text{SiO}_2(\text{s}) \rightarrow \text{SiO}_2(\text{g})$ at the average experimental temperature. Similarly, for liquid SiO_2 , $\Delta H^* = 133.9 \pm 10.0$ and $\Delta S^* = 37.4 \pm 4.4$, similar to $\Delta H^\circ = 136.0 \pm 8.0$ and $\Delta S^\circ = 37.8 \pm 0.6$ (ref. 13) for the reaction $\text{SiO}_2(\text{l}) \rightarrow \text{SiO}_2(\text{g})$. These correspondences imply that solid and liquid SiO_2 evaporate not as $\text{SiO}(\text{g})$, but as $\text{SiO}_2(\text{g})$.

A similar explanation cannot be invoked to understand the small evaporation coefficient for $\text{MgO}(\text{s})$, however, because J_e for the oxidized magnesium gas species (J_e in Fig. 1b) is much smaller than J_v (which would make $\alpha_v > 1$, violating the requirement that α_v must be less than unity). Therefore periclase probably evaporates according to the decomposition reaction $\text{MgO}(\text{s}) \rightarrow \text{Mg}(\text{g}) + \text{O}(\text{g})$ and/or $\text{O}_2(\text{g})$, with a significant excess energy barrier involved.

Solid forsterite was found to evaporate stoichiometrically at the experimental temperatures (1,600–1,800 °C). (This is also the case at equilibrium: thermodynamic calculations from literature data indicate that the vapour produced from forsterite should have an overall composition equal to forsterite stoichiometry. This has been demonstrated by experiments performed near equilibrium^{5,14}.) Molten forsterite also evaporates almost stoichiometrically. Only the most highly evaporated residues contained a trace (~5 vol%) of SiO_2 -rich glass (~65 wt% SiO_2). Because the evaporation of both solid and liquid forsterite is stoichiometric (or almost so), α_v values for the MgO and SiO_2 components are equal, and were found to be 0.09–0.16 at temperatures beneath the melting point of forsterite (1,600–1,890 °C) and 0.20–0.21 above it (1,890–2,050 °C).

As in the case of SiO_2 evaporation discussed above, however, these results depend on values of J_e calculated assuming that the vapour species are present in equilibrium proportions. If J_e is recalculated (J_e in Fig. 1c) on the assumption that the vapour species produced when Mg_2SiO_4 evaporates are $\text{SiO}_2(\text{g})$, $\text{Mg}(\text{g})$, and $\text{O}(\text{g})$ (or $\text{O}_2(\text{g})$), consistent with the analysis described above of the evaporation of pure $\text{SiO}_2(\text{s})$ and 1) and $\text{MgO}(\text{s})$, α_v is found to be 0.4–0.6 (or 0.3–0.5, for $\text{O}_2(\text{g})$ instead of $\text{O}(\text{g})$) for solid forsterite and 0.8 (or 0.6) for liquid forsterite, closer to unity than the values cited above. The fact that these α_v values for forsterite are smaller than those for solid and liquid SiO_2 (~1) is interpreted to mean that the evaporation of the MgO component of forsterite is subject to a kinetic effect similar to but smaller than that observed for pure MgO .

Although the possibility that $\text{SiO}(\text{g})$ is the main evaporative Si species cannot be eliminated on the basis of the above discussion, there is another reason to expect $\text{SiO}_2(\text{g})$ to have this role. Oxides and tellurides of group IVA and VA elements such as CO_2 , GeO_2 , NO_2 , P_2O_5 , As_2O_3 , PbTe and SnTe , the chemical bonds of which are covalent, as are the Si–O bonds in SiO_2 and Mg_2SiO_4 , have been shown to evaporate mostly into

molecules with the same molecular formula (or their polymers) as in the condensed phases^{15–18}. Na_2CO_3 decomposes on evaporation into Na, CO_2 and O_2 molecules¹⁹; $\text{Na}(\text{g})$ and $\text{CO}_2(\text{g})$ can be seen as analogues of $\text{Mg}(\text{g})$ and $\text{SiO}_2(\text{g})$.

Thermodynamic calculations of the type used here to obtain J_e and also Knudsen-cell mass spectrometry studies²⁰ show that $\text{SiO}(\text{g})$ is the principal silicon gas species in equilibrium with condensed SiO_2 and Mg_2SiO_4 . This means that $\text{SiO}_2(\text{g})$ generated by evaporation must subsequently decompose into $\text{SiO}(\text{g})$ and O (or O_2). A subsequent gas-phase reaction of this sort does not, however, affect the mechanism or rate of evaporation. In a restricted volume, such as inside a Knudsen cell, equilibrium will be reached between species through the reaction $\text{SiO}_2(\text{g}) \rightleftharpoons \text{SiO}(\text{g}) + \text{O}(\text{g})$ (or $\frac{1}{2}\text{O}_2$). If the vapour consists of molecules produced only from condensed silica and forsterite, $\text{SiO}(\text{g})$ and an equivalent amount of $\text{O}(\text{g}) + \text{O}_2(\text{g})$ outnumber $\text{SiO}_2(\text{g})$ in the gas phase at equilibrium. Thus if a molecular stream through a small orifice of the cell is introduced into a mass spectrometer, $\text{SiO}(\text{g})$ rather than $\text{SiO}_2(\text{g})$ will be found to be the predominant gas species.

The range of temperatures used in my experiments may be higher than the temperatures at which evaporation and condensation occurred in the solar nebula. As there is a linear relationship between $\log \alpha_v$ and $1/T$, it is easy to extrapolate my results to lower temperatures, where α_v decreases. For forsterite, α_v is ~0.12 at 1,700 °C, ~0.04 at 1,400 °C. Thus, at 1,400 °C forsterite evaporates ~25 times more slowly than it would if α_v were unity; it is more 'refractory' than another material would be that has the same equilibrium vapour pressure but a larger α_v .

Consequently, in a nebular situation forsterite might evaporate at the same rate as some mineral that is intrinsically much more refractory (one with a much smaller equilibrium vapour pressure) for which the evaporation coefficient is near unity. If evaporation were interrupted, both minerals would still be found in the residue, making it seem that they are equally refractory. In fact, preliminary studies²¹ of the evaporation kinetics of refractory planetary components such as the rare-earth oxides, TiO_2 , CaO , Al_2O_3 , CaAl_4O_7 and MgAl_2O_4 have shown that these materials do have larger α_v values than that of forsterite. Therefore, although forsterite is much more volatile than these oxides under equilibrium conditions, it can behave as if it were equally refractory during a non-equilibrium evaporation event.

Probably the most important aspect of this study is that it introduces a timescale into considerations of two fundamental physical processes, evaporation and condensation, in the primordial solar nebula, which cannot be known from equilibrium considerations as previously developed. □

Received 8 February; accepted 19 July 1990.

1. Urey, H. C. *The Planets* (Yale University Press, New Haven, 1952).
2. Anders, E. *Phil. Trans. R. Soc. A* **285**, 23–40 (1977).
3. Lord, H. C. *Icarus* **4**, 279–288 (1965).
4. Grossman, L. *Geochim. cosmochim. Acta* **36**, 597–619 (1972).
5. Mysen, B. O. & Kushiro, I. *Am. Miner.* **73**, 1–19 (1988).
6. Wood, J. A. *A. Rev. Earth planet. Sci.* **16**, 53–72 (1988).
7. Hirth, J. P. & Pound, G. M. *Condensation and Evaporation* (Pergamon, Oxford, 1963).
8. Langmuir, I. *Phys. Rev.* **2**, 329–342 (1913).
9. Somorjai, G. A. & Lester, J. E. *Prog. Solid State Chem.* **4**, (ed. Reiss, H.) 1–53 (Pergamon, Oxford, 1967).
10. Sasamoto, T., Lee, H. L. & Sata, T. *Yogyo-Kyokai-Shi* **82**, 603–610 (1974).
11. Firsova, L. P. & Nesmeyanov, An. N. *Soviet J. phys. Chem.* **34**, 1279–1281 (1960).
12. Nagai, S., Niwa, K., Shinmei, M. & Yokokawa, T. *JCS Faraday Trans. I* **69**, 1628–1634 (1973).
13. Chase, M. W. et al. *JANAF Thermochemical Tables*, 3rd edn (Dow Chemical Company, Midland, 1986).
14. Nagahara, H., Kushiro, I., Mysen, B. O. & Mori, H. *Nature* **331**, 516–518 (1988).
15. Sasamoto, T., Kobayashi, M. & Sata, T. *Shitsuryo Bunseki* **29**, 249–255 (1981).
16. Muenow, D. W., Uy, O. M. & Margrave, J. L. *J. inorg. nucl. Chem.* **32**, 3459–3467 (1970).
17. Brittain, R. D., Lau, K. H. & Hildenbrand, D. L. *J. phys. Chem.* **86**, 5072–5075 (1982).
18. Northrop, D. A. *J. phys. Chem.* **75**, 118–132 (1971).
19. Motzfeldt, K. *J. phys. Chem.* **59**, 139–147 (1955).
20. Porter, R. F., Chupka, W. A. & Inghram, M. G. *J. chem. Phys.* **23**, 216–217 (1955).
21. Hashimoto, A., Holmberg, B. B. & Wood, J. A. *Meteoritics* **24**, 276 (1989).

ACKNOWLEDGEMENTS. I thank J. Wood for discussions and B. Holmberg for electron microprobe analyses.

Isotropic cellular automaton for modelling excitable media

Mario Markus & Benno Hess

Max-Planck-Institut für Ernährungsphysiologie, Rheinlanddamm 201, D-4600 Dortmund 1, FRG

EXCITABLE media, exemplified by the chemical system of the Belousov-Zhabotinsky (BZ) reaction, are often modelled theoretically through the integration of sets of partial differential equations that describe their dynamics. An alternative approach is to use cellular automata¹⁻⁴, which sacrifice insight into the detailed physical mechanisms for the benefit of being able to reproduce the observed patterns of behaviour at low computational cost. But the cellular automata used in previous work have been anisotropic (mainly square or hexagonal), leading to the problem that this anisotropy tends to be propagated into the patterns produced. Here we describe a means of generating isotropic cellular automata, which are able to reproduce a wide range of observed modes of behaviour ranging from spiral-type BZ patterns to structures reminiscent of turbulence.

Various cellular automata for modelling fluid dynamics have been proposed, including a highly anisotropic one based on square cells⁵, a two-dimensional isotropic one based on a hexagonal lattice⁶ and a three-dimensional isotropic one that operates in a four-dimensional space⁷. Despite the complexity of this last model, strategies have been developed for efficient calculation⁸. In excitable media^{9,10}, however, all of the cellular automata so far proposed are anisotropic because the shape of the cells of the periodic lattice propagates to macroscopic scales. This leads to waves having the shape of the cells (such as squares², hexagons¹ and cubes³) or to cell-shaped waves with rounded corners⁴. Here we use a cellular automaton having randomized space discretization. This automaton yields isotropic wave propagation in two and three dimensions. Of course, there are systems, especially in physiology, in which anisotropy appears in nature. For the consideration of such anisotropy—our model must be modified to include the specific properties of the medium.

An excitable medium is a dynamic system capable of being 'excited' by an above-threshold perturbation. After excitation, it becomes 'refractory', slowly returning to the 'receptive' state. In the receptive state, or close to it, it can be excited again. Such a medium can support waves without attenuation. Examples of such waves are electrical activity propagating along an axonal membrane¹¹, contraction of heart muscle¹², spreading depression on the cerebral cortex around epileptic foci¹³ and on the retina¹⁴, contractions on the surface of fertilized eggs¹⁵, waves of reactants in the BZ and other chemical reactions^{1,16-23}, waves of cyclic adenosine 3',5'-monophosphate in social amoeba²⁴, infectious diseases in biological populations^{25,26} and waves of shock-driven star formation in spiral galaxies^{1,27}.

The simpler excitable media are characterized by equations of the type $\partial A/\partial t = f(A, B) + D_A \nabla^2 A$, $\partial B/\partial t = g(A, B) + D_B \nabla^2 B$ (see, for example, refs 20, 28). Typical nullclines $f(A, B) = 0$ and $g(A, B) = 0$, as well as trajectories are shown in Fig. 1. We make time discrete as follows: the transition from the receptive state (α) or from a refractory state close to α (such as δ) to the excited state (β) requires one iteration. The reverse transition usually takes much longer (see, for example, refs 29, 30) and is given by $n+1$ iterations, where n is a model parameter.

Figure 2 shows the geometry of our cellular automaton. A state S is assigned to each point: $S = 0$ (receptive state), $S = n+1$ (excited state) and $S = n, n-1, \dots, 2, 1$ (refractory states). At each point, S decreases by 1 at each iteration step if no excitation takes place at that point. A parameter $S_{\max}$ is introduced, having the following meaning: excitation is not possible for $S > S_{\max}$

(for example, point γ in Fig. 1), but it is possible for $0 \leq S \leq S_{\max}$ (for example, point δ in Fig. 1). The excitation at a point is triggered by the points in its neighbourhood (circle of radius R). A point becomes excited if the number ν of excited points inside the circular neighbourhood is larger than a given threshold m , analogous to ref. 2. We make the assumption $m = m_0 + pS$ ($p \geq 0$, $0 \leq S \leq S_{\max}$), which takes into consideration that the triggering threshold for the excitation of a point is smaller the closer the point is to the receptive state.

Before determining a state $S(t+\tau)$ from a state $S(t)$ (τ is the time step per iteration), we determine an intermediate state $\sigma(t)$ which thereafter yields $S(t+\tau)$ by local averaging⁴. The intermediate state is determined by the following rules: (1) for excitation, if $0 \leq S(t) \leq S_{\max}$ and $\nu \geq m_0 + pS(t)$, then $\sigma(t) = n+1$; (2) for stationarity of the receptive state, if $S(t) = 0$ and $\nu < m_0$, then $\sigma(t) = 0$; (3) for loss of excitation and of refractoriness in time, $\sigma(t) = S(t) - 1$ for any case not considered before.

The new state $S(t+\tau)$ is obtained by the closest integer to the average of $\sigma(t)$ over all points inside the circular neighbourhood. (This takes into account transport processes, as in ref. 4). We did not perform this averaging for $\sigma = 0$, $\sigma = 1$ or $\sigma = n+1$, because it leads to numerical instabilities in wavefronts and

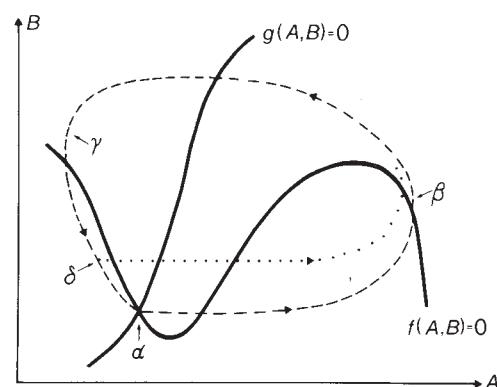


FIG. 1 Typical phase-plane diagram for an excitable medium, showing the local dynamics of an excitation variable A and a recovery variable B . α , receptive state; β , excited state; γ , non-excitable refractory state; δ , excitable refractory state. Solid curves, nullclines; dashed curve, trajectory starting from the receptive state; dotted curve, trajectory starting from an excitable refractory state.

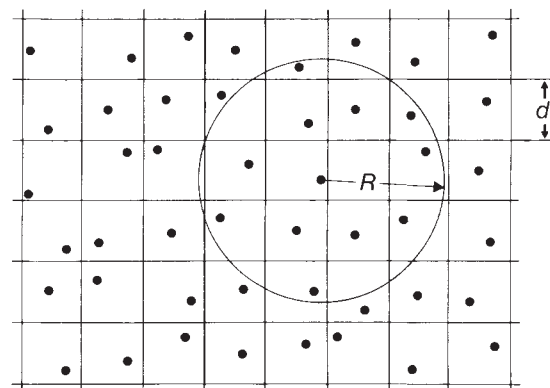


FIG. 2 Geometry of the cellular automaton. The two-dimensional (or three-dimensional) medium is divided into square (or cubic) cells of side d . One point is placed randomly in each cell at $t=0$ and this distribution remains the same throughout the simulation. This constraint of one point per cell, instead of a random distribution of each point within the whole medium, prevents large inhomogeneities while keeping the number of points low. The neighbourhood of each point is defined by a circle with radius R .

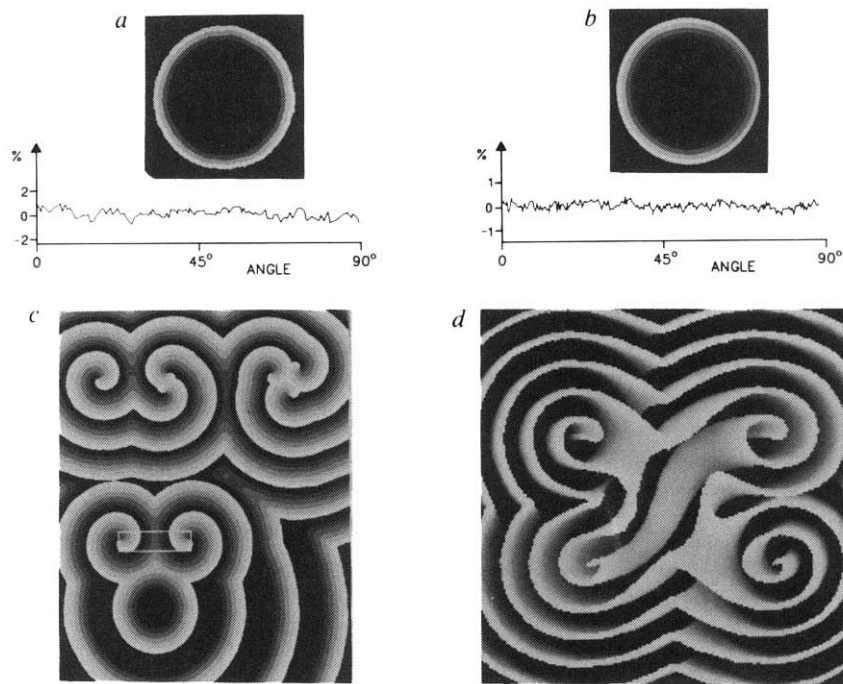


FIG. 3 *a, b*, Two-dimensional waves generated by an excited circle, demonstrating the isotropy of the model ($m_0=1$, $n=3$, $p=S_{\max}=0$, $t=15\tau$). The percent deviation from the mean of the distance travelled by the excited wavefront is shown below as a function of the angle. (Because of the symmetry of the automaton, we plot this as the average over the four quadrants.) *a*, 200×200 cells, $r=6$. *b*, 400×400 cells, $r=12$. *c*, Two-dimensional waves ($r=10$, $m_0=1$, $p=S_{\max}=0$, $t=68\tau$, 750×600 cells). In the rectangular domain n is set to 16 for $10 \leq t \leq 26$. At other times and elsewhere, $n=4$. The period of the pacemaker point at the lower left is 10τ . Grey levels change from black to white as the state S changes from 0 to $n+1$. The initial conditions are a broken plane wave for the one-armed spiral (upper left) and two and three broken plane waves coming from different directions and sharing cross-sections of the excited fronts for two- and three-armed spirals (upper middle and upper right). *d*, Three-dimensional simulation (pair of linked scroll rings each with unit twist). $r=3$, $n=3$, $m_0=1$, $p=S_{\max}=0$, $t=24\tau$, $200 \times 200 \times 200$ cells. Only the excited state in the back half of the medium (100 cells deep) is shown. Darkness of the grey levels increases as cells lie deeper in the cubic lattice.

wavebacks. Altogether, we have five control parameters: $r=R/d$, n , $S_{\max}$, m_0 and p . In approximations where only the receptive state can be excited, $p=S_{\max}=0$, and there are only three parameters. If we place the points at the same positions in all cells (instead of randomly), or assume a square instead of a circular neighbourhood, isotropy is lost. If we do not use local averaging, the tip of a one-armed spiral (defined by the point where the wavefront curvature changes sign) moves along a straight line segment, in contradiction to experiments²¹.

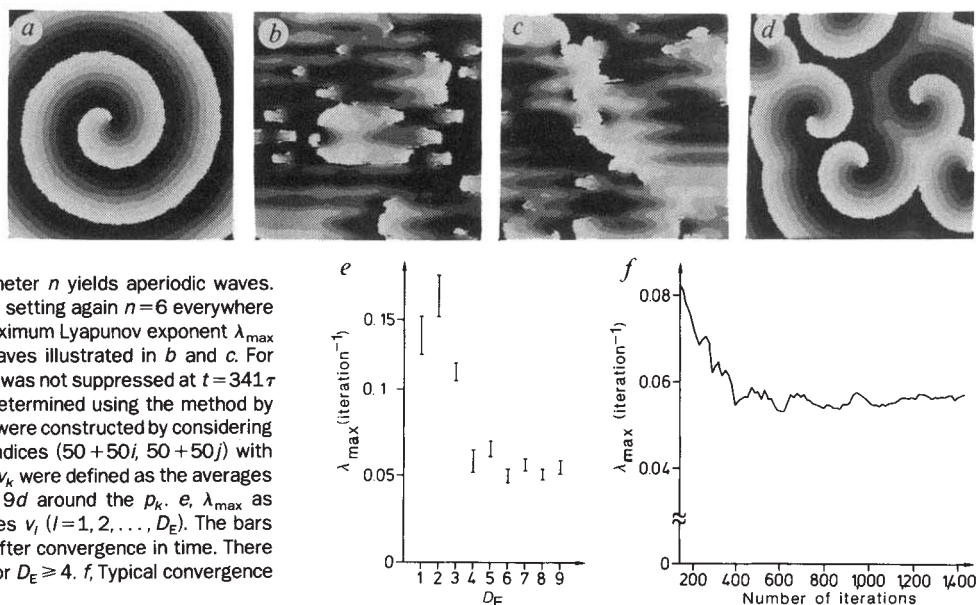
Figure 3*a, b* illustrates the isotropy of the automaton, showing circular waves at two different resolutions. Figure 2*c* shows one-, two- and three-armed spirals similar to the BZ reaction^{17,21}. In particular, the tip of a one-armed spiral rotates around a circular region (spiral core^{20,21}). Further, the distance between the spiral tips of two- and three-armed spirals oscillates in time, as observed in experiments¹⁷.

At the bottom left of Fig. 3*c*, we show circular waves generated by a periodic pacemaker circle, analogous to the sinoatrial node of the heart. A transient inhomogeneity (in the rectangular domain) induces a pair of spirals. Because of their higher frequency, the spirals annihilate the circular waves. This simulates the sudden increase of heart rate owing to inhomogeneities in the myocardium¹². Figure 3*d* shows a simulation in three dimensions, namely a pair of twisted scroll rings, which have been simulated by partial differential equations²⁸ and by anisotropic automata³, but have not yet been observed experimentally. Simpler scroll waves have been postulated from observations in thick heart muscle¹² and have been observed in the BZ reaction, both in solution^{16,18} and in gels²².

After choosing a grid length d , physical counterparts of the model parameters can be established as follows: m_0 is the number of points within a circle of radius R_{crit} , this being the

FIG. 4 Simulated transition from periodicity (*a*) to aperiodicity (*b, c*) and back to periodicity (*d*) for $r=6$, $m_0=1$, $p=S_{\max}=0$, 200×200 cells. In *a* ($t=0$) $n=6$ everywhere. During the time interval $0 \leq t \leq 341$ (including *b* at $t=73\tau$ and *c* at $t=340\tau$) n depends on the ordinate, varying between 4 and 20 in a piecewise linear manner with period $20d$ and equal absolute values of the slopes of all linear segments.

The spatial periodicity of the control parameter n yields aperiodic waves. Reorganization into spirals takes place after setting again $n=6$ everywhere for $t \geq 341\tau$ (Fig. 3*d* is at $t=389\tau$). *e, f*, Maximum Lyapunov exponent $\lambda_{\max}$ for the characterization of the aperiodic waves illustrated in *b* and *c*. For the calculation of $\lambda_{\max}$ the spatial periodicity was not suppressed at $t=341\tau$ but maintained until $t=1,400\tau$. $\lambda_{\max}$ was determined using the method by Wolf *et al.*³¹. For this purpose, phase spaces were constructed by considering 9 cells p_k ($k=1, \dots, 9$) given by the grid indices $(50+50i, 50+50j)$ with $i=0, 1, 2$ and $j=0, 1, 2$. The phase variables v_k were defined as the averages of the states S inside circles with radius $9d$ around the p_k . *e*, $\lambda_{\max}$ as determined in phase spaces with coordinates v_i ($i=1, 2, \dots, D_E$). The bars indicate the range in which $\lambda_{\max}$ oscillates after convergence in time. There is no significant dependence of $\lambda_{\max}$ on D_E for $D_E \geq 4$. *f*, Typical convergence of $\lambda_{\max}$ ($D_E=8$).



(measurable) radius of the smallest excited circle capable of initiating a wave; R is the radius of a circle having the following properties: if the centre of the circle is at a distance $c\tau$ from a planar wavefront (c being the velocity of free planar waves and τ the interaction time, that is, the transition time from the receptive to the excited state), then there are m_0 points at the intersection of the circle with the excited part of the planar wavefront; $(n+1)\tau$ is the refractory time, or the time for relaxation from the excited to the receptive state; $(n+1-S_{\max})\tau$ is the time after excitation for which no further excitation is possible. Finally, p describes, to a first approximation, the rate of growth of the distance between the refractory trajectory and the separatrix in the phase plane beyond which excitation takes place.

Experiments on the BZ reaction²³ can be simulated by setting $r=5$, $n=9$, $m_0=2$, $p=2$ and $S_{\max}=8$. Measurements give spiral period $T=17.3$ s, spiral velocity $v=76\text{ }\mu\text{m s}^{-1}$, velocity of plane waves $c=82.5\text{ }\mu\text{m s}^{-1}$, $R_{\text{crit}}=23.1\text{ }\mu\text{m}$ and diffusion coefficient $D_a=2\times 10^{-5}\text{ cm}^2\text{ s}^{-1}$. Simulations yield $T=10.5\tau$ and $v=3.8d/\tau$, leading to the relations $d=33\text{ }\mu\text{m}$ and $\tau=1.65$ s. These relationships allow us to calculate $c=85\text{ }\mu\text{m s}^{-1}$, $R_{\text{crit}}=26.4\text{ }\mu\text{m}$ and $D_a=2.24\times 10^{-5}\text{ cm}^2\text{ s}^{-1}$. (D_a was obtained as the slope of the plot of normal wavefront velocity against curvature K , for $K\rightarrow 0$.) These results are in good agreement with experiments. (Non-integer lengths are possible in the simulations because the averages are calculated over the randomly distributed points).

Simulations of plane waves by periodically exciting the edge of the medium yield a dispersion relation analogous to theoretical predictions and experiments (see ref. 19). With the parameters mentioned above we obtain a monotonic increase of c from $57\text{ }\mu\text{m s}^{-1}$ at a wave period $T=13$ s to $85\text{ }\mu\text{m s}^{-1}$ at $T\geq 18$ s.

A spatially periodic refractory time—described by a spatially periodic parameter n in our automaton—disrupts periodic waves (Fig. 4a) yielding aperiodic behaviour (Fig. 4b, c). Suppression of this spatial periodicity of n yields reorganization into periodic waves (Fig. 4d). The positive sign of the maximum Lyapunov exponent $\lambda_{\max}$ (see ref. 31), given in Fig. 4e, f for large D_E and large t , is one of the features of deterministic chaos. (In contrast, the time evolutions of the waves shown in Fig. 4a, d are characterized by $\lambda_{\max}=0$.) It is thus expected that excitable media, such as those containing living cells, may display chaotic behaviour for particular spatial variations of the control conditions. □

Received 23 March; accepted 12 July 1990.

- Madore, B. F. & Freedman, W. L. *Science* **222**, 615–616 (1983).
- Zykov, V. S. & Mikhailov, A. S. *Soviet Phys. Dokl.* **31**, 51–52 (1988).
- Winfree, A. T., Winfree, E. M. & Seifert, H. *Physica D* **17**, 109–115 (1985).
- Gerhardt, M. & Schuster, H. *Physica D* **36**, 209–221 (1989).
- Hardy, J., Pomeau, Y. & de Pazzis, O. *J. math. Phys.* **14**, 1746–1759 (1973).
- Frisch, U., Hasselbacher, B. & Pomeau, Y. *Phys. Rev. Lett.* **56**, 1505–1508 (1986).
- d'Humières, D., Lallemand, P. & Frisch, U. *Europhys. Lett.* **2**, 291–297 (1986).
- Rivet, J. P., Hénon, M., Frisch, U. & d'Humières, D. *Europhys. Lett.* **7**, 231–236 (1988).
- Zykov, V. S. *Simulations of Wave Processes in Excitable Media* (Manchester University Press, 1988).
- Holden, A. V., Markus, M. & Othmer, H. G. (eds) *Nonlinear Wave Processes in Excitable Media* (Plenum, London, in the press).
- Hodgkin, A. L. & Huxley, A. F. *J. Physiol.* **117**, 500–544 (1952).
- Medvinsky, A. B., Panfilov, A. V. & Pertsov, A. M. in *Self-Organization* (ed. Krinsky, V. I.) 195–199 (Springer, Berlin, 1984).
- Koroleva, V. I. & Bures, J. *Brain Res.* **173**, 209–215 (1979).
- Gorelova, N. A. & Bures, J. *J. Neurobiol.* **14**, 353–363 (1983).
- Hara, K., Tydeman, P. & Kirschner, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 462–466 (1980).
- Winfree, A. T. *Science* **181**, 937–939 (1973).
- Agladze, K. I. & Krinsky, V. I. *Nature* **296**, 424–426 (1982).
- Welsh, B. J., Gomati, J. & Burgess, A. E. *Nature* **304**, 611–614 (1983).
- Sevcikova, H. & Marek, M. *Physica D* **39**, 15–21 (1989).
- Keener, J. P. & Tyson, J. J. *Physica D* **21**, 307–324 (1986).
- Müller, S. C., Plesser, Th. & Hess, B. *Physica D* **24**, 71–86 (1987).
- Jahrke, W., Henze, C. & Winfree, A. T. *Nature* **336**, 662–665 (1988).
- Forster, P., Müller, S. C. & Hess, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6831–6834 (1989).
- Tomchik, K. J. & Devreotes, P. N. *Science* **212**, 443–446 (1981).
- Carey, A. B., Giles, R. H. Jr & McLean, R. G. *Am. J. trop. Med. Hyg.* **27**, 573–580 (1978).
- Murray, J. D., Stanley, E. A. & Brown, D. L. *Proc. R. Soc. B* **229**, 111–150 (1986).
- Schulman, L. S. & Seiden, P. E. *Science* **233**, 425–431 (1986).
- Nandapurkar, P. J. & Winfree, A. T. *Physica D* **29**, 69–83 (1987); **D35**, 277–288 (1989).
- Field, R. J. & Noyes, R. M. *J. chem. Phys.* **60**, 1877–1884 (1974).
- Martiel, J.-L. & Goldbeter, A. *Biophys. J.* **52**, 807–828 (1987).
- Wolf, A., Swift, J. B., Swinney, H. G. & Vastano, J. A. *Physica D* **16**, 285–317 (1985).

ACKNOWLEDGEMENTS. This work was supported by the Stiftung Volkswagenwerk, Hanover.

Two-isotope characterization of N_2O in the Pacific Ocean and constraints on its origin in deep water

Kyung-Ryul Kim* & H. Craig

Isotope Laboratory, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093, USA

THREE decades after the first study of nitrous oxide (N_2O) in the ocean¹, the marine geochemistry of this compound is of exceptional interest for two reasons: first, the need to understand the role of N_2O in both nitrification and denitrification in the sea^{2–5}, and second, the importance of N_2O in the stratospheric ozone cycle⁶ and as a greenhouse trace gas that, molecule for molecule, is 200 times more effective than carbon dioxide⁷. Related questions concern the role of the ocean as a source or sink of tropospheric N_2O and the processes responsible for its production and consumption in the sea. Our approach to these problems is one that has been successful in studies of H_2O geochemistry⁸ and sources of atmospheric CO (ref. 9): the simultaneous study of two isotopic components of a molecule, here the $^{15}\text{N}/^{14}\text{N}$ and $^{18}\text{O}/^{16}\text{O}$ ratios. We have measured these ratios in N_2O in three Pacific Ocean profiles and in Pacific marine air. Relative to atmospheric N_2O , dissolved N_2O is slightly depleted in both heavy isotopes down to a depth of ~ 600 m, but in deep and bottom water, N_2O is everywhere enriched in ^{15}N and ^{18}O . Thus, contrary to recent predictions^{10,11}, N_2O production in deep water is characterized by heavy-isotope enrichment in both nitrogen and oxygen. These enrichments may be due to microbial reduction of N_2O produced by nitrification, or they may reflect a prior ^{15}N , ^{18}O enrichment in an intermediate compound (for example, NH_2OH) from which the light isotopes are preferentially channelled into nitrite and nitrate during nitrification.

Our profiles were made at three of the Pacific GEOSECS stations¹²: station 314 in the South Pacific gyre, station 204 in the North Pacific gyre and station 347, the Northeast Pacific 'GEOSECS I' station¹³. Seawater samples of total volume $\sim 2,000$ litres were collected at each depth and stripped of gas in a 'Macbeth Machine' developed for ^{39}Ar tracer studies¹⁴ (the primary purpose of these long stations). Collection of each profile required about a week. In the laboratory, N_2O was separated and converted to N_2 and CO_2 for measurements of the $^{15}\text{N}/^{14}\text{N}$ and $^{18}\text{O}/^{16}\text{O}$ ratios. The results are given in Table 1 as δ_{atm} values relative to atmospheric N_2 and O_2 isotope ratios, together with other data from the three stations. Table 1 also shows our measurements on N_2O in North Pacific and South Pacific marine air: these data are very similar to published values for N_2O in North Pacific air¹⁰— $\delta^{15}\text{N}_{\text{atm}}=6.5\text{--}8.6\%$ (average 7.2%)—and for Albany, New York air¹⁵— $\delta^{18}\text{O}_{\text{atm}}=18.7\text{--}24.2\%$ (average 21.4%).

Figure 1 shows the two-isotope characteristics of N_2O in the ocean and marine atmosphere. The seawater data from the South Pacific station GS 314 lie on a linear trend (slope = 0.30) that brackets mean atmospheric N_2O ($\delta^{15}\text{N}\approx 7.0$, $\delta^{18}\text{O}\approx 20.7$): the near-surface samples (0–600 m) are slightly depleted in both ^{15}N and ^{18}O relative to atmospheric N_2O , whereas the deeper samples are enriched in the heavier isotopes. Samples from the North Pacific stations are even more enriched in the heavy isotopes, and for the most part form a band with a much smaller slope (~ 0.10) at $\delta^{18}\text{O}$ values $> 27\%$. These samples correspond generally to dissolved NO_3^- and O_2 concentrations > 35 and $< 160\text{ }\mu\text{M kg}^{-1}$.

Figure 2 shows the depth profiles of $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ in N_2O

* Present address: Department of Oceanography, Seoul National University, Seoul, 151-742, Korea.

TABLE 1 Nitrogen and oxygen isotope ratios in marine and atmospheric N₂O

Seawater N ₂ O				N ₂ O*			
Station	Location	Depth (m)	NO ₃ (μM kg ⁻¹)	O ₂	δ ¹⁵ N (‰)	δ ¹⁸ O (‰)	δ ¹⁸ O(O ₂)† (‰)
GS314	23°44' S	300	3	204	6.7	18.0	4.6
	153°37' W	400	7	204	6.4	17.9	4.6
	(S. Pacific)	575	23	200	6.4	18.2	5.0
		800	30	192	7.6	21.9	5.4
		2,400	36	149	8.9	26.5	5.8
		3,750	34	180	8.7	25.2	5.4
		4,900	32	195	8.4	24.2	4.8
GS204	31°22' N	200	9	220	6.3	—	3.6
	150°02' W	450	21	180	6.9	21.6	5.0
	(N. Pacific)	1,050	43	13	8.6	31.7	12.0
		2,200	40	84	9.7	32.3	10.0
		3,500	36	140	9.1	28.2	7.0
		5,000	35	159	8.8	28.7	6.0
GS347	28°30' N	250	27	112	8.6	21.3	8.2
	121°29' W	400	36	37	9.8	28.4	9.4
	(N.E. Pacific)	500	39	17	8.5	27.5	9.7
		600	40	8	8.6	31.7	11.0
		1,000	44	24	8.8	29.3	12.0
		1,700	42	60	9.2	33.0	10.0
		2,800	39	113	9.4	30.4	8.0
Marine air N ₂ O							
Position		Date		δ ¹⁵ N	δ ¹⁸ O		
24°26' S	154°19' W	4/13/83		7.1	19.6		
12°37' S	141°40' W	4/21/83		7.1	20.4		
01°15' N	124°32' W	5/2/83		7.0	21.4		
18°05' N	122°51' W	5/7/83		7.0	21.2		
30°58' N	158°25' W	8/85		6.9	20.8		
Average				7.0	20.7		

Seawater samples were collected in Gerard barrels at the sites of GEOSECS stations 204, 314 and 347: we collected 2,000 litres of water at each depth, in four casts of two barrels each. NO₃⁻ and dissolved O₂ data tabulated are from the GEOSECS data at these stations¹². Dissolved gases were stripped out quantitatively aboard ship using a 60-l flow-through Macbeth Machine, dried at 0 °C, and returned in SpectraSeal-treated aluminium cylinders. Air samples were collected by compressing ~2,000 litres of air into similar cylinders with a Rix clean-air compressor. In the laboratory, N₂O and CO₂ were trapped at liquid-nitrogen temperature, CO₂ was removed by passage through Ascarite and N₂O was then further purified by gas chromatography separation. N₂O was converted to N₂ and CO₂ by reaction at 700 °C with carbon rods wrapped with platinum wire. After 7 min of reaction, CO₂ and the remaining N₂O were frozen in liquid nitrogen, and N₂ was toepled into a sample tube. The N₂O and CO₂ were then warmed up and reacted again: after three such cycles the conversion was generally complete, as measured by the absence of further N₂ production. N₂ was purified of residual CO by passage through I₂O₅ (ref. 9) and CO₂ and N₂ were then analysed for ¹⁸O/¹⁶O and ¹⁵N/¹⁴N ratios on the 6" radius mass spectrometer SAMSON.

* The isotopic data are δ_{atm} values (relative to atmospheric N₂ and O₂): δ(‰) = [(R/R⁺) - 1] × 10³, where R⁺ is the ¹⁵N/¹⁴N or ¹⁸O/¹⁶O ratio in atmospheric N₂ or O₂ respectively. Analyses of commercial-tank N₂O used as a substandard averaged δ¹⁵N = 2.5 ± 0.16 and δ¹⁸O = 14.1 ± 0.36.

† Values of δ¹⁸O_{atm}(‰) of dissolved O₂ relative to atmospheric O₂, estimated from the δ¹⁸O-dissolved O₂ relationship in the Pacific¹⁸. Atmospheric O₂ is +23.5‰ [=δ¹⁸O_{SMOW}(‰)] relative to Standard Mean Ocean Water²⁷. Values of δ¹⁸O relative to SMOW are converted to δ¹⁸O_{atm} by the relation δ¹⁸O_{atm} = -23.0 + δ¹⁸O_{SMOW}/1.0235.

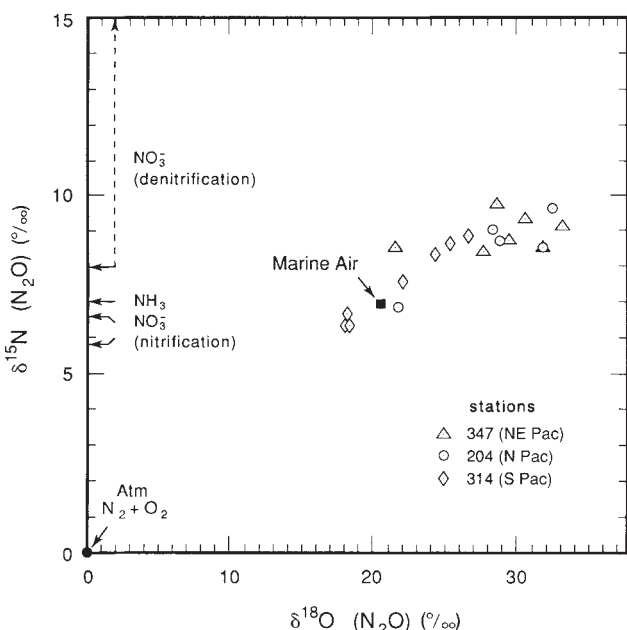


FIG. 1 ¹⁵N/¹⁴N and ¹⁸O/¹⁶O variations in Pacific Ocean N₂O. δ_{atm} is the ‰ enrichment of each ratio relative to atmospheric N₂ or O₂. δ¹⁸O_{atm} in dissolved O₂ associated with these samples ranges from 4 to 12‰ because of preferential bacterial consumption of ¹⁶O relative to ¹⁸O. δ¹⁵N values for oceanic ammonia²⁸ and nitrate^{24,28} on the ordinate show the range of δ¹⁵N observed in nitrification and in regions of denitrification. In the eastern tropical Pacific, δ¹⁵N enrichments up to 18‰ are observed in residual NO₃⁻ because of the more rapid reduction of ¹⁴NO₃⁻. Marine atmospheric N₂O is bracketed by the shallow- and deep-water isotopic values at the South Pacific station, perhaps indicating a strong oceanic source for tropospheric N₂O. Continental N₂O is similar to marine in both δ¹⁵N (ref. 29) and δ¹⁸O (ref. 15) values, although with a larger range. A sample of soil N₂O, however, was observed¹⁵ to have δ¹⁸O_{atm} = +0.5‰ (much lighter than atmospheric N₂O), raising a question as to the importance of soil N₂O in the tropospheric budget because the isotopic composition of N₂O in near-surface sea water is close to that of the atmosphere. Measurements of both isotopes in soil N₂O will resolve this important question.

at stations 204 and 314, together with nitrate and dissolved oxygen from the same locations¹². In the South Pacific (station 314), the deep-water isotopic maxima are associated with the very broad extrema in NO_3^- and O_2 at about the same depths and are clearly related to the biological processes of NO_3^- production and O_2 consumption in these deep waters. In the North Pacific (station 204), however, the deep-water enrichments are considerably greater, with maxima that lie well below the NO_3^- and O_2 extrema at $\sim 1,000$ m. Here the NO_3^- and O_2 concentration extrema have shoaled in response to concentration changes in the Intermediate Water¹⁶, whereas the N_2O isotopic data are largely unaffected.

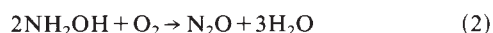
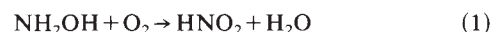
N_2O production in deep water is generally thought to be due to oxidative regeneration of nitrate (nitrification), as N_2O is correlated negatively with dissolved O_2 and positively with NO_3^- throughout the water column²⁻⁵. It has been proposed that, because of kinetic effects, biological nitrification produces isotopically light N_2O , depleted both in ^{15}N (refs 10, 11) and in ^{18}O (ref. 17). Our results, however, show that deep-ocean N_2O is enriched in both ^{15}N and ^{18}O relative to atmospheric N_2O . This can be seen clearly when the isotopic data are plotted as a function of nitrate concentration, the characteristic parameter for nitrification in the ocean (Fig. 3). At low NO_3^- concentrations the ^{15}N and ^{18}O enrichments do decrease slightly below the atmospheric values, but they increase rapidly in all three profiles as nitrate rises above $30 \mu\text{M kg}^{-1}$, showing that N_2O associated with nitrification is quite generally enriched in both heavy isotopes relative to atmospheric N_2O .

The ^{18}O enrichment in N_2O is due in part to ^{18}O enrichment of the dissolved O_2 in deep water¹⁸, because some, if not all, of the oxygen in N_2O must be derived from this O_2 (ref. 19). In Fig. 3 we show the $\delta^{18}\text{O}$ values for dissolved oxygen at the same depths as the N_2O samples, as derived from the $\delta^{18}\text{O}$ - O_2 relationships in the region of each station¹⁸. Even though $\delta^{18}\text{O}$ of the dissolved O_2 increases with nitrate concentration, the $^{18}\text{O}/^{16}\text{O}$ enrichment in N_2O is much larger than in the dissolved O_2 . Relative to the baseline minimum enrichments in near-surface waters (<600 m), $\delta^{18}\text{O}$ of N_2O increases by 15% to the maximum observed values (33%), in contrast to the 7% increase in $\delta^{18}\text{O}$ of dissolved oxygen over the same range.

Heavy-isotope enrichment in microbial systems is almost always due to kinetic fractionation effects, which favour more rapid utilization of lighter isotopes in consumption processes (such as respiration), resulting in a pool of residual molecules enriched in heavy isotopes (for example, dissolved oxygen¹⁸). When, as in the case of N_2O throughout most of the ocean (or O_2 in the euphotic zone), there is a net production rather than consumption, then for heavy-isotope enrichment in the com-

ponent, there must in general be simultaneously occurring production and consumption, with an appropriately larger kinetic fractionation in the consumption process depending on the ratio of consumption to production. The delineation of these processes for N_2O then depends on whether nitrification or denitrification is the process responsible for the production of N_2O . Thus Yoshida *et al.*²⁰ reported $\text{N}_2\text{O}/\text{NO}_3^-$ ratios in the upper few kilometres of water at two North Pacific stations that were 5–10 times higher than the production ratios measured in bottles with a ^{15}N tracer. They proposed that N_2O in these waters is primarily produced by anaerobic denitrification in the interiors of organic particles, and simultaneously consumed by a separate denitrification process in the water column. The actual $\text{N}_2\text{O}/\text{NO}_3^-$ production ratio in the water column ($\sim 10^{-3}$) is, however, roughly that expected from observed ratios of N_2O supersaturation to O_2 undersaturation in regions of nitrification^{4,5,21}. We also note that the possibility of increased NO_3^- production by the acceleration of O_2 consumption in bottles needs to be evaluated in these tracer experiments. For these reasons, and because the model of Yoshida *et al.* requires both production and consumption of N_2O by denitrification (with the former dominating), we conclude that the generally accepted model²⁻⁵ of N_2O production by nitrification is more probably correct.

Production by nitrification requires a coupled consumption process with preferential removal of the light isotopes to account for the general heavy-isotope enrichment in deep water. The first stage of nitrification is the microbial oxidation of NH_4^+ by dissolved O_2 to hydroxylamine. The process then continues by two channels^{22,23} for the oxidation of hydroxylamine:



in which reaction (1) is dominant. Several possibilities exist for light-isotope selection in the required consumption step. One is simply microbial reduction of N_2O occurring simultaneously with nitrification, perhaps by denitrification in the interiors of organic particulates. O. C. Zafriou (personal communication) has suggested what is perhaps a more probable mechanism: isotopic fractionation in the oxidation of hydroxylamine to nitrite in reaction (1). Because of the dominance of this channel, the more rapid oxidation rates of the light isotopes will result in a pool of heavy-isotope-enriched NH_2OH , from which the deep-water N_2O is drawn with a corresponding (though lesser) heavy-isotope enrichment.

Nitrification seems to be the most plausible origin for N_2O in deep water from the viewpoints of material balance and ^{15}N

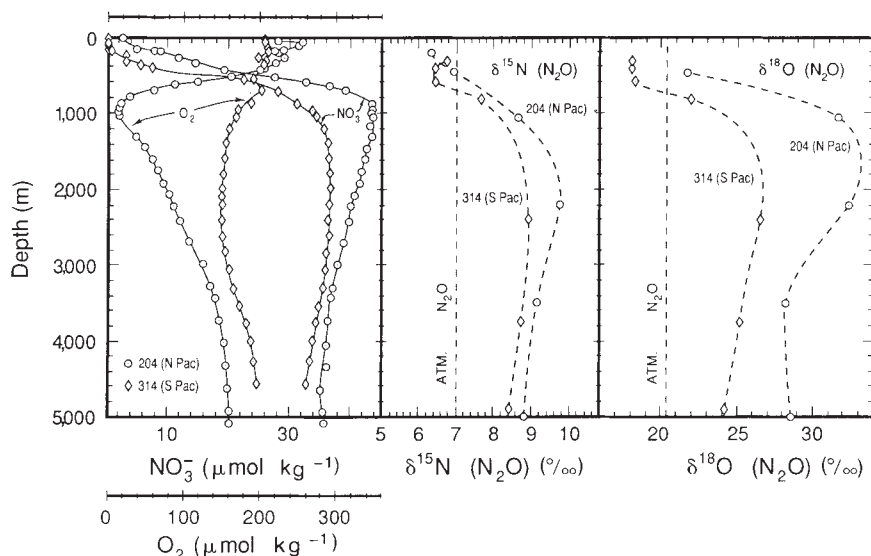


FIG. 2 $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ profiles in N_2O in two Pacific Ocean stations. (Profiles at station 347, not shown, are similar to station 204.) The NO_3^- and dissolved O_2 profiles from the same locations are shown in the left panel. Note the large ^{15}N and ^{18}O enrichments for both N and O in deep water below ~ 600 m. Both heavy-isotope enrichments increase from the South to the North Pacific, paralleling the increase in NO_3^- and decrease in O_2 that results from the 'deep metabolism'¹⁶. The scarcity of isotopic data reflects the fact that each data point represents 2,000 kg of water: gases were totally extracted from ~ 40 tons of water during this work.

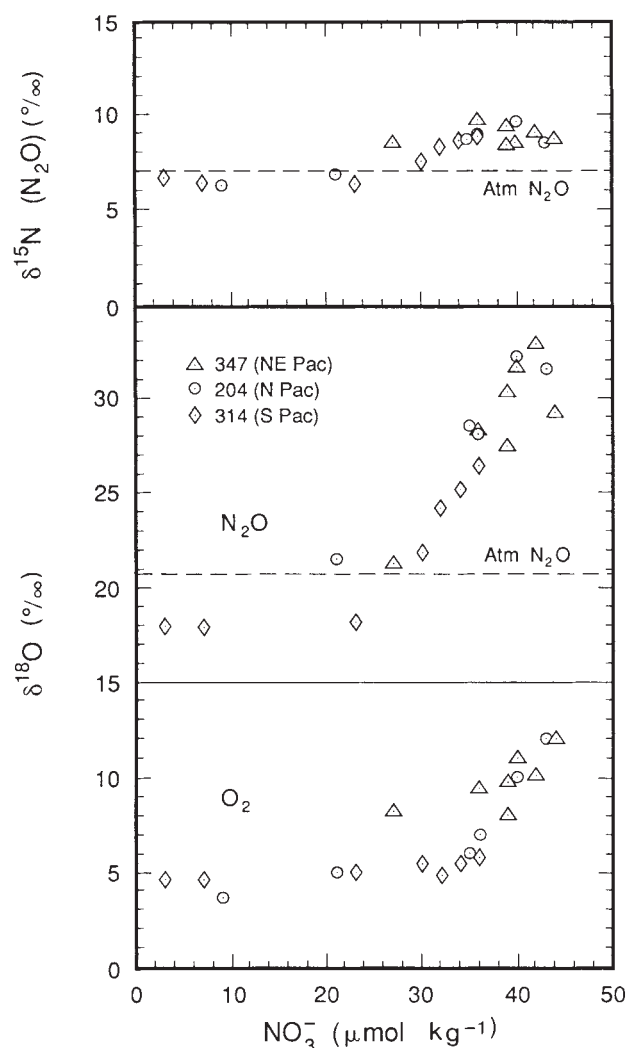


FIG. 3 $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ as a function of dissolved nitrate, the characteristic parameter for nitrification. The isotopic enrichments of ^{18}O in dissolved O_2 from the same depths and areas¹⁸ are shown in the lower panel. At low NO_3^- concentrations, in depths $< \sim 600$ m, $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ are less than the atmospheric values (compare Fig. 1), probably because N_2O production exceeds consumption in the early stages of nitrification. At NO_3^- levels $> 30 \mu\text{M kg}^{-1}$, the ^{18}O enrichments in both the increasing N_2O and decreasing O_2 concentrations increase approximately linearly with NO_3^- , but the increase in $\delta^{18}\text{O}$ of the N_2O is about twice that of the dissolved O_2 (15% compared with 8%). Thus the fractionation effect in dissolved O_2 accounts for about half the ^{18}O enrichment in the deep-water N_2O .

and ^{18}O enrichments. But the Eastern Equatorial Pacific, an area of active denitrification and upwelling, may be the major site of N_2O flux from ocean to atmosphere^{4,24-26}. The two-isotope characterization of N_2O in this region will be the next major step in understanding the sources of this trace gas in the atmosphere. □

15. Wahlen, M. & Yoshinari, T. *Nature* **313**, 780-782 (1985).
16. Craig, H. *J. geophys. Res.* **76**, 5078-5086 (1971).
17. Yoshinari, T. & Wahlen, M. *Nature* **317**, 349-350 (1985).
18. Kroopnick, P. & Craig, H. *Earth planet. Sci. Lett.* **32**, 375-388 (1976).
19. Hollocher, T. C., Tate, M. E. & Nicholas, D. J. D. *J. biol. Chem.* **256**, 10834-10836 (1981).
20. Yoshida, N. *et al. Nature* **342**, 895-897 (1989).
21. Kaplan, W. A. & Wofsy, S. C. in *Advances in Aquatic Microbiology* Vol. 3, (eds Jannasch, H. W. & Williams, P. J.) 181-206 (Academic, London, 1985).
22. Fenchel, T. & Blackburn, T. H. *Bacteria and Mineral Cycling* (Academic, London, 1979).
23. Butler, J., Elkins, J. W., Thompson, T. M. & Egan, K. B. *J. geophys. Res.* **94**, 14865-14877 (1989).
24. Cline, J. D. & Kaplan, I. R. *Mar. Chem.* **3**, 271-299 (1975).
25. Weiss, R. F. *Eos* **62**, 894 (1981).
26. Cline, J. D., Wisegarver, D. P. & Kelly-Hansen, K. *Deep Sea Res.* **34**, 857-873 (1987).
27. Kroopnick, P. & Craig, H. *Science* **175**, 54-55 (1972).
28. Miyake, Y. & Wada, E. *Records Oceanogr. Works Japan* **9**, 37-53 (1967).
29. Yoshida, N. & Matsuo, S. *Geochim. J.* **17**, 231-239 (1983).

ACKNOWLEDGEMENTS. We thank Y. Horibe for helpful assistance in the laboratory and for discussions, R. Comer for running our many Gerard casts and air collections, and V. Craig for help at sea and ashore. We are grateful to O. C. Zafriou for critical discussions. This work was supported by the Chemical Oceanography Program of the NSF.

No evidence from multichannel reflection data for a crustal magma chamber in the MARK area on the Mid-Atlantic Ridge

Robert S. Detrick*, John C. Mutter†, Peter Buhl‡ & Isaac I. Kim‡

* Graduate School of Oceanography, University of Rhode Island, Kingston, Rhode Island 02881, USA

† Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York 10964, USA

‡ Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA

THE role of crustal magma chambers in the construction of oceanic crust along the slowly spreading Mid-Atlantic Ridge (MAR) is still highly disputed¹. Some investigators have argued for small, short-lived crustal magma bodies², whereas others have postulated the presence of permanent reservoirs centred between ridge offsets, which expand and contract in response to a fluctuating magma supply³. Thermal modelling^{4,5}, detailed seismic refraction experiments at several different locations along the MAR⁶⁻⁹, and micro-earthquake^{10,11} and teleseismic-source mechanism studies¹² preclude the existence of large, steady-state magma chambers under the entire MAR rift valley; however, smaller, localized crustal magma bodies could be present, at least intermittently. Here we use multichannel seismic reflection profiles to look for evidence of a crustal magma chamber along the MAR south of the Kane fracture zone, in one of the most magmatically robust and hydro-thermally active areas presently known along the MAR. Although some sub-basement reflectors are imaged, we do not see a shallow, large-amplitude, intracrustal reflector that might be associated with a crustal magma body. If a crustal magma chamber was associated with the most recent volcanic episode along this part of the MAR, our results suggest that it was extremely short-lived.

Seismic reflection techniques have proven to be a useful tool for imaging crustal magma chambers along fast spreading ridges¹³⁻¹⁷ and across a back-arc spreading centre¹⁸. Along the northern East Pacific Rise (EPR), for example, a high-amplitude, flat-lying intracrustal event is observed 1-2 km below the sea floor on both multichannel¹³⁻¹⁶ and single-channel¹⁷ seismic reflection profiles. This event can be unequivocally tied to the top of a seismic low-velocity zone and is associated with the large drop in compressional and shear-wave velocities consistent with the presence of molten or partially molten material^{19,20}. If similar crustal magma bodies exist, even intermittently, beneath portions of the MAR rift valley, seismic reflection techniques should be capable of detecting them.

Received 19 December 1989; accepted 24 July 1990.

1. Craig, H. & Gordon, L. I. *Geochim. cosmochim. Acta* **27**, 949-955 (1963).
2. Yoshinari, T. *Mar. Chem.* **4**, 189-202 (1976).
3. Cohen, Y. & Gordon, L. I. *Deep Sea Res.* **25**, 509-524 (1978).
4. Elkins, J. W., Wofsy, S. C., McElroy, M. B., Kolb, C. E. & Kaplan, W. A. *Nature* **275**, 602-606 (1978).
5. Cohen, Y. & Gordon, L. I. *J. geophys. Res.* **84**, 347-353 (1979).
6. Crutzen, P. J. *Q. J. R. met. Soc.* **96**, 320-325 (1970).
7. Laci, A. A., Hansen, J., Lee, P., Mitchell, T. & Lebedeff, S. *Geophys. Res. Lett.* **8**, 1035-1038 (1981).
8. Craig, H. *Science* **133**, 1702-1703 (1961).
9. Stevens, C. M. *et al. Earth planet. Sci. Lett.* **16**, 147-165 (1972).
10. Yoshida, N., Hattori, A., Saino, T., Matsuo, S. & Wada, E. *Nature* **307**, 442-444 (1984).
11. Yoshida, N. *Nature* **335**, 528-529 (1988).
12. Broecker, W. S., Spencer, D. W. & Craig, H. *Geosecs Vol. 3: Pacific Expedition Hydrographic Data* (National Science Foundation, Washington, DC, 1982).
13. Craig, H. & Weiss, R. F. *J. geophys. Res.* **75**, 7641-7647 (1970).
14. Craig, H., Kim, K., Loosli, H. H., Weidman, U. & Kalt, P. *Eos* **64**, 736 (1983).

In January 1989 we obtained ~700 km of multichannel seismic reflection data in the MARK area, a well studied portion of the MAR south of the Kane fracture zone near 23° N (Fig. 1). The MAR in this area consists of at least two independent spreading cells separated by a discordant zone or small-offset fracture zone near 23°15' N^{21,22}. In the northern part of the rift valley, immediately south of the Kane transform, recent volcanic activity has constructed a 500-m-high, 40-km-long ridge on the rift valley floor^{22,23}, the largest feature of its kind ever reported for the MAR²¹. Although the age of the lavas forming this large neovolcanic ridge is not known, the lack of alteration, the thin ferro-manganese rinds, and the light dusting of pelagic sediment all indicate that this feature is very young, perhaps only a few thousand years old²¹. A large, active hydrothermal vent field, known as the Snake Pit hydrothermal area, is located on the crest of this ridge, ~30 km south of the Kane transform^{24,25}. Quartz geobarometry on fluid samples from the Snake Pit area indicate that a high-temperature reaction zone, possibly associated with a magma chamber, is present 1–2 km below the sea floor²⁶. A refraction profile shot ~5 km east of this neovolcanic ridge found a relatively normal, but slightly thin, crustal section underlying the eastern half of the rift valley at this latitude⁸, but provided no constraints on the possible existence of crustal magma bodies beneath the neovolcanic ridge or the Snake Pit hydrothermal area.

Among the multichannel lines obtained in our survey were profiles across the rift valley at the latitude of the Snake Pit hydrothermal area (line 838 in Fig. 1) and along the neovolcanic ridge from the Kane transform south to ~23° N (line 841). These lines were shot with an 87.6-l (5,346 in³) airgun array and received on a 96-channel digital streamer with 25-m-long active sections. This system has been used successfully to map numerous intracrustal reflectors in Mesozoic crust in the western North Atlantic²⁷. Stacking velocities were based on crustal-

velocity models derived from Purdy and Detrick's⁸ rift-valley refraction line. Coherent noise produced by scattering and diffraction from the rough MAR topography in and out of the plane of the section is a severe problem in these data. We were partially successful in suppressing this noise by applying a prestack frequency-wavenumber ($f-k$) filter to shot gathers²⁸ and attenuating-ray parameters (p) $> \pm 0.35$ s km⁻¹. A typical processing sequence thus consisted of bandpass filtering (6–30 Hz), $f-k$ filtering, bandpass filtering, spherical divergence correction, gathering into common mid-points, normal moveout (NMO) correction, stacking, bandpass filtering and, optionally, post-stack $f-k$ migration.

The processed sections presented in Fig. 2 show several weak, discontinuous crustal reflectors. A subhorizontal event is present at ~7-s two-way travel time along the middle portion of line 838 and on the southern half of line 841. This event occurs close to, but slightly below, the expected depth of the crust/mantle boundary based on the refraction results of Purdy and Detrick⁸ and may be a Moho reflection. Several other weak events are also present on line 841, most notably the dipping reflectors between 6- and 7-s two-way time from CDP 400 to 750 and a shallower subhorizontal event at CDP 850 near the Snake Pit hydrothermal area (thin arrows in Fig. 2). A large-amplitude intracrustal reflector, like that observed at the east Pacific Rise^{13–15} or across the Valu Fa Ridge¹⁸, is not present beneath the Snake Pit hydrothermal area on either line 838 or 841.

To investigate the possible origin of the events seen on line 841 and the geological significance of the absence of a 'magma chamber' reflection in these data, we have modelled seafloor scattering and transmission using an approach developed by Kim and Orcutt²⁹. The synthetic data were calculated from a Sea Beam bathymetry map³⁰ gridded at a 200-m interval along an ~15-km-wide swath centred on the profile. The scattering algorithm uses a triangular representation of the sea floor. The

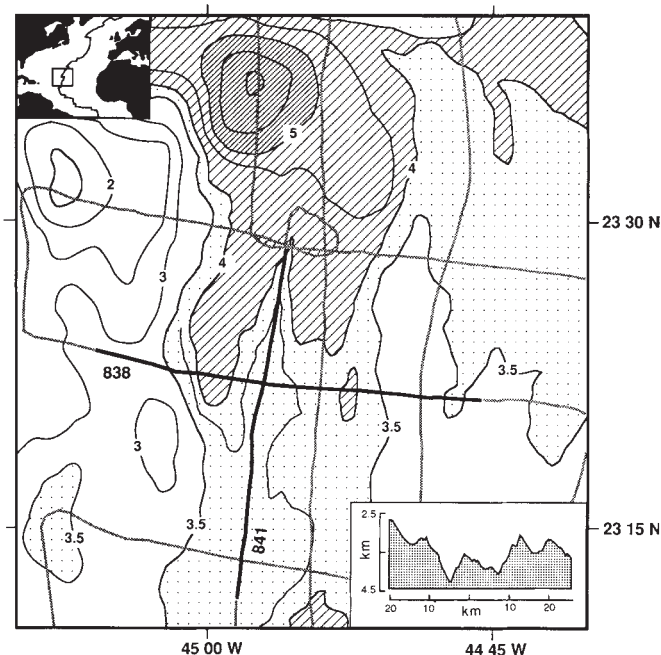


FIG. 1 Bathymetry map of the northern part of the MARK rift valley just south of its intersection with the Kane fracture zone (500-m contour interval; depths greater than 3,500 m are shaded). The inset presents a bathymetric cross-section of the rift valley at 23° 20' N showing the neovolcanic ridge that dominates this part of the rift valley. The solid lines show the portions of profiles 838 and 841 presented in Fig. 2. Line 841 runs along the crest of the neovolcanic ridge; line 838 crosses the ridge at the location of the Snake Pit hydrothermal area^{24,25}.

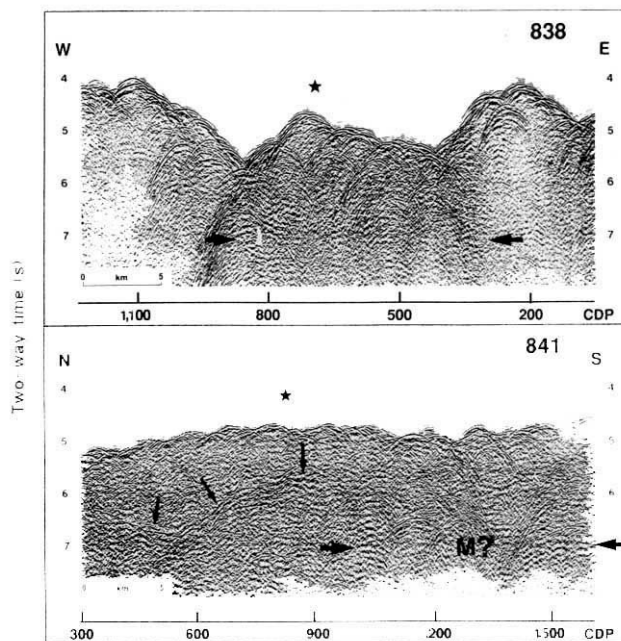


FIG. 2 Stacked time sections for seismic profiles 838 (top) and 841 (bottom); see Fig. 1 for location. The Snake Pit hydrothermal area is indicated by a star. A weak subhorizontal event, a possible Moho reflection, is present at ~7-s two-way travel time on both profiles (thick arrows). Several other weak events are also present on line 841, most notably the dipping events between 6- and 7-s two-way time from common depth point (CDP) 400 to 750 and a shallower subhorizontal event at CDP 850 near the Snake Pit hydrothermal area (thin arrows). A large-amplitude intracrustal reflector which might be associated with a crustal magma body is not present beneath the neovolcanic ridge on either profile.

analytically calculated diffraction response from each triangular element is summed to produce the scattered response^{31,32}. Figure 3a,b compares a portion of the stacked section from line 841 with a zero-offset, single-channel synthetic section composed only of scattering from the sea floor. This section reproduces many of the stronger events seen on line 841. In particular, it demonstrates that the events seen between CDPs 400 and 850 (arrows in Fig. 3a) are out-of-plane diffractions from the axial

trough that borders the neovolcanic ridge on the west (see Fig. 1) and are not real intracrustal reflections.

Could the scattering effects owing to the rugged topography in the MAR rift valley completely mask a reflection from a crustal magma chamber? To address this question we have calculated a second synthetic section in which a simple interface with an impedance contrast comparable to that expected for a boundary between crustal rocks ($\sim 6.1 \text{ km s}^{-1}$) and molten lava ($< 3.0 \text{ km s}^{-1}$) was assumed to exist at a depth of 5.5 km along the entire length of line 841. This model, based on the results of Harding *et al.*¹⁹, corresponds to a normal-incidence reflection coefficient of 0.33. The amplitude of this event assuming a flat sea floor was calculated using the reflectivity method, then adjusted for the effect of a rough sea floor using a Kirchhoff-Helmholtz transmission algorithm²⁹. This result was then added to the scattering synthetic calculated above to create the section displayed in Fig. 3c. This section demonstrates that large-scale scattering and wavefront distortion caused by MAR topography with spatial wavelengths of a few hundred metres or more is not strong enough to mask a reflection from the roof of an essentially molten crustal body, provided it is laterally continuous on the scale of several kilometres. If the reflection coefficient is substantially smaller, or if the magma chamber is laterally discontinuous, the reflector will be weaker and could be indistinguishable from the scattered energy. Based on our experience with seismic reflection data from the EPR^{15,16}, bodies with widths as narrow as 100–200 m can be resolved on multi-channel reflection profiles provided they are associated with a large enough (0.2–0.4) reflection coefficient. If the reflection coefficient is substantially smaller, even a laterally continuous body will be hard to observe. Modelling such as that shown in Fig. 3c indicates that if the reflection coefficient is ≤ 0.07 –0.14 the large amount of scattered energy along line 841 in the MARK area can completely obscure such an event. A reflection coefficient this small, however, would imply either a largely solidified magma body (V_p , compressional wave velocity, $> 5.0 \text{ km s}^{-1}$) or a much more gradual transition from solid to molten rock than has been inferred from other ridge-crest seismic studies^{19,20}.

These scattering synthetic sections demonstrate that the absence of a large-amplitude intracrustal reflector in the MARK area is not solely a consequence of the rougher seafloor topography in the median valley of the MAR. Our results indicate that a predominantly molten crustal magma chamber of significant dimensions does not presently exist beneath the Snake Pit hydrothermal area or elsewhere along the neovolcanic ridge in the northern part of the MARK rift valley. This result is particularly significant in that this is one of the most magmatically robust and hydrothermally active areas known on the MAR. The absence of a crustal magma chamber here, where one would most expect to find one, suggests that, along slow-spreading ridges, magma chambers are extremely short-lived, if they exist at all. □

Received 28 May; accepted 17 July 1990.

1. Macdonald, K. C. A. *Rev. Earth planet. Sci.* **10**, 155–190 (1982).
2. Nisbet, E. G. & Fowler, C. M. R. *Geophys. J. R. astr. Soc.* **54**, 631–660 (1978).
3. Stakes, D., Shervais, J. & Hopson, C. *J. geophys. Res.* **89**, 6995–7028 (1984).
4. Sleep, N. H. *J. geophys. Res.* **80**, 4037–4042 (1975).
5. Kuznir, N. J. & Bott, M. H. P. *Geophys. J. R. astr. Soc.* **47**, 83–95 (1976).
6. Fowler, C. M. R. *Geophys. J. R. astr. Soc.* **47**, 459–491 (1976).
7. Fowler, C. M. R. *Geophys. J. R. astr. Soc.* **54**, 167–183 (1978).
8. Purdy, G. M. & Detrick, R. S. *J. geophys. Res.* **91**, 3739–3762 (1986).
9. Loudon, K. E., White, R. S., Potts, G. & Forsyth, D. W. *J. geol. Soc. Lond.* **143**, 795–806 (1986).
10. Lilwall, R. C., Francis, T. J. G. & Porter, I. T. *Geophys. J. R. astr. Soc.* **51**, 357–369 (1977).
11. Toomey, D. R., Solomon, S. C. & Purdy, G. M. *J. geophys. Res.* **93**, 9093–9112 (1988).
12. Huang, P. Y. & Solomon, S. C. *J. geophys. Res.* **93**, 13445–13477 (1988).
13. Herron, T. J., Stoffa, P. L. & Buhl, P. *Geophys. Res. Lett.* **7**, 989–992 (1980).
14. Hale, L. D., Morton, C. J. & Sleep, N. H. *J. geophys. Res.* **87**, 7707–7717 (1982).
15. Detrick, R. S. *et al. Nature* **326**, 35–41 (1987).
16. Mutter, J. C. *et al. Nature* **336**, 156–158 (1988).
17. Avedik, F. & Geli, L. *Geology* **15**, 857–860 (1987).
18. Morton, J. L. & Sleep, N. *Circum-Pacific Council for Energy and Miner. Resour. Sci. Ser.* **2**, 441–453 (1985).
19. Harding, A. J. *et al. J. geophys. Res.* **94**, 12163–12196 (1989).

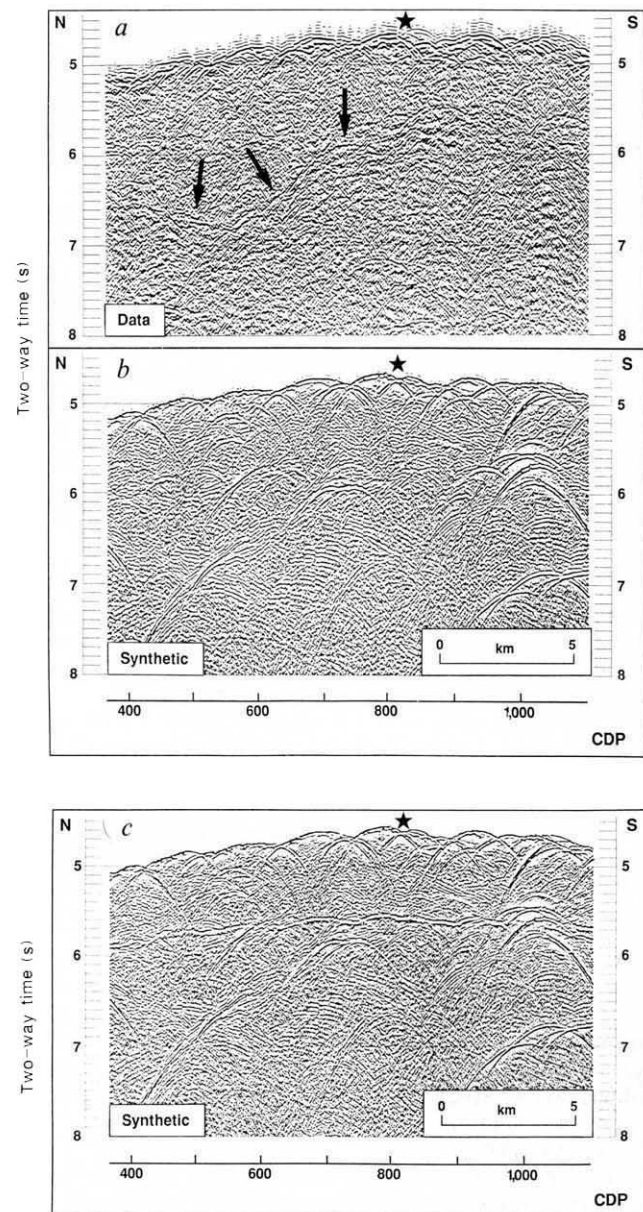


FIG. 3 a, Stacked time section for a portion of line 841 along the neovolcanic ridge near the Snake Pit hydrothermal area (star). b, A zero-offset, single-channel synthetic section for the northern half of line 841 composed of only scattering from the seafloor topography in an ~ 15 -km-wide swath centred on the profile. This section reproduces many of the scattered events seen on line 841 and demonstrates that the events seen between CDPs 400 and 850 (arrows) are out-of-plane diffractions from the axial trough that borders the median valley ridge on the west (see Fig. 1), not real intracrustal reflections. c, same as b, except that an interface with a reflection coefficient comparable to that expected for a molten magma chamber is assumed to exist at a depth of 5.5 km along line 841. This synthetic section demonstrates that large-scale scattering and wavefront distortion caused by MAR topography with spatial wavelengths of a few hundred metres or more is not strong enough to mask a reflection from the roof of a molten crustal body, provided it is laterally continuous on this scale.

20. Vera, E. E. *et al.* *J. geophys. Res.* (in the press).
21. Karson, J. A. *et al.* *Nature* **328**, 681–685 (1987).
22. Kong, L. S. L., Detrick, R. S., Fox, P. J., Mayer, L. A. & Ryan, W. B. F. *Mar. geophys. Res.* **10**, 59–90 (1988).
23. Pockalny, R. A., Detrick, R. S. & Fox, P. J. *J. geophys. Res.* **94**, 3179–3193 (1988).
24. ODP Leg 106 Scientific Party *Geology* **14**, 1004–1007 (1986).
25. Karson, J. & Brown, J. *Mar. geophys. Res.* **10**, 91–107 (1988).
26. Campbell, A. C. *et al.* *Nature* **335**, 514–518 (1988).
27. White, R. S. *et al.* *Geology* **18**, 462–465 (1990).
28. Lerner, K., Chambers, R., Yang, M., Lynn, W. & Wai, W. *Geophysics* **48**, 854–886 (1983).
29. Kim, I. I. & Orcutt, J. A. *Eos* **70**, 1317 (1989).
30. Detrick, R. S. *et al.* *Eos* **65**, 1006 (1984).
31. Trorey, A. W. *Geophysics* **35**, 762–784 (1970).
32. Hilterman, F. J. *Geophysics* **47**, 784–808 (1982).

ACKNOWLEDGEMENTS. We thank the officers and crew of the RV *Robert C. Conrad* for their assistance, J. Alsop, R. Forman, E. Morris and J. Dolan for help with the data analysis, and K. Macdonald for comments. This work was supported by the Joint Oceanographic Institutions/US Science Advisory Committee and the NSF.

Elastic strength of the Slave craton at 1.9 Gyr and implications for the thermal evolution of the continents

John Grotzinger & Leigh Royden

Department of Earth, Atmospheric and Planetary Sciences,
Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA

THE secular cooling of the Earth since its formation, together with the law of radioactive decay, imply that the Earth's internal heat generation and heat loss must have been two to three times higher in Archaean and Proterozoic times than today, but this inference is difficult to test directly. Here we present a flexural analysis of the Proterozoic Killohigok basin, in the Archaean Slave craton of Canada, which shows that 1.9 Gyr ago the craton had an effective elastic thickness of only 12 ± 4 km. Because the present effective elastic thickness of the region is $\sim 100 \pm 25$ km, these results imply a two- to fourfold decrease in the thermal gradient beneath the Slave craton since 1.9 Gyr, and indicate that at 1.9 Gyr—700 Myr after the accretion of the continental lithosphere forming the craton—the thick continental root beneath this part of the Canadian shield had not yet formed.

Previous estimates of geothermal gradients in the early Earth have been based on the petrology of ancient igneous and metamorphic rocks^{1,2} or on the height of ancient mountain belts as inferred from palaeopressure estimates³. These estimates are applicable only to areas of magmatic or orogenic activity, offering no constraints on the early thermal structure of stable cratons or ancient oceanic lithosphere. An alternative means of estimating palaeotemperature structures, which is applicable to cratonic lithosphere, involves measurement of ancient flexural rigidities (or equivalently, effective elastic plate thicknesses). The effective elastic plate thickness of modern oceanic lithosphere has been shown to be roughly equivalent to the depth to the 450 °C (± 150 °C) isotherm⁴. In regions of continental crust this relationship seems to be complicated by ductile failure of the lower continental crust at sufficiently high temperatures, but the elastic plate thickness probably corresponds to the depth of the ~ 450 °C (± 150 °C) isotherm for cold lithosphere and to the depth between the Moho and the 450 °C (± 150 °C) isotherm for hot oceanic lithosphere.

The early Proterozoic Killohigok Basin (~ 1.9 Gyr old), located on Archaean continental crust of the Canadian shield, affords an excellent opportunity to examine the flexural behaviour and thermal structure of cratonic lithosphere in early Proterozoic time (Fig. 1). This foreland basin was formed by flexure of the continental lithosphere during coeval convergence and thrusting along an adjacent active margin. The basin sediments preserve

a record of basin geometry, thus allowing for direct analysis of the flexural behaviour of the underlying lithosphere in early Proterozoic time. By comparing the elastic plate strength of this cratonic lithosphere in early Proterozoic time with that of modern lithosphere of the Canadian shield and elsewhere, one can obtain a rough comparison between continental geothermal gradients in the early Proterozoic relative to those that exist today.

The Killohigok Basin is located on continental crust along the northeastern margin of the Archaean Slave craton, and is bounded to the west by Archaean basement rocks of the Slave craton and to the east by the Thelon orogenic belt⁵. Sedimentation within the basin was contemporaneous with thrusting and convergence within the Thelon orogenic belt, which contains high-grade metamorphic and igneous rocks dated at 2,010–1,920 Myr (refs 6–8). Subsequent warping and erosion has exposed a continuous cross-section of the original basin for nearly 350 km across strike. The basin is offset by one transcurrent fault with 115 km of sinistral slip, but the motion can be removed and the fault restored to ± 5 km (ref. 8).

Reconstruction of the original basin geometry from these stratigraphic data reveals an originally asymmetric basin thickening eastwards and with a maximum thickness of more than 5 km in the deepest parts^{5,9}. Facies interpretations, palaeobathymetric estimates and determination of time lines have been discussed in detail elsewhere^{5,10}. The stratigraphically lowest sequence in the basin includes an initial passive-margin terrace of shallow (<5 m water depth) marine sediments (the Kimerot Platform) which is everywhere <500 m thick. The platform represents a coastal plain wedge on the most cratonic part of the passive margin⁵. The lack of significant subsidence of the Kimerot Platform during cooling of the rifted margin indicates that little extension or heating of the lithosphere occurred beneath the platform, as does the lack of normal faults beneath the platform. During the initial foredeep basin development, the passive margin underwent drowning of its outer edge and uplift within its interior. In the deeper parts of the basin the overlying sediments are deep-water (>500 m water depth) submarine fan and interfan deposits; units deposited at the same time in shallower parts of the basin are shallow shelf sands and shales. These are overlain by shallow (<50 m water depth) marine shelf deposits (the Link formation) and finally by a wedge of southeastward-thickening synorogenic fluvial coarse sandstones and conglomerates deposited on a northwest-sloping braidplain⁹. The sequences analysed here were deposited between about 1,967 ± 5 Myr and 1,921 ± 10 Myr ago^{10,11}.

To study the flexural strength of the lithosphere beneath the Killohigok Basin, we assume that the lithosphere behaves as a thin elastic sheet overlying a fluid asthenosphere. The net deflection of the basement between deposition of any two isochronous sedimentary horizons can thus be described by the equation governing flexure of an elastic sheet under applied loads. The net vertical deflection of the basement between deposition of any two sedimentary horizons corresponds to the thickness of sediment deposited in the intervening interval, corrected for (1) sediment compaction, (2) the water depth (relative to sea level) in which the two sediments were deposited, and (3) eustatic sea level changes (Table 1). By choosing as our reference horizons the top of the shallow (<5 m water depth) marine Kimerot Platform and the top of the shallow (<5 m water depth) marine lower Burnside formation, we can eliminate (2) because deposition of these two sediment horizons occurred close to sea level. Sediment decompaction was accomplished by assuming subsequent burial beneath ~ 3 km of non-marine molasse and other overlying units⁹, and standard values for initial porosities of sandstone (25%) and shale (60%). Individual units were decompacted according to the thicknesses and sand/shale ratios give by Grotzinger *et al.*¹⁰.

We used an established method¹² to find the theoretical solution to the elastic plate equation (including change in eustatic

TABLE 1 Kilohigok Basin depths

Stratigraphic section	Present thickness (m)	Decompacted thickness (m)
1	840	1,115
2	210	355
3	245	400
4	300	485
5	730	985
6	1,020	1,325
7	2,170	3,540
8	3,960	5,355

sea level) that provides the best fit to the observed deflection of the Slave craton for a fixed flexural rigidity. (To obtain the best-fit flexural solution, sea level was allowed to vary but by no more than 300 m). This procedure was repeated for a range of flexural rigidities until upper and lower bounds on the flexural rigidity of the lithosphere could be established. Figure 2 shows that acceptable fits between theoretical and observed deflection profiles exist for flexural rigidities D between 4×10^{21} and 3×10^{22} N m, thus constraining the effective elastic plate thickness to between 8 and 16 km at ~ 1.9 Gyr. Lithospheric flexure and basin development were accompanied by a rise in eustatic sea level of at least 250 m, consistent with observations that the entire Slave craton was covered by a sedimentary blanket during the same time interval. It is difficult to establish an upper limit on the eustatic rise in sea level that caused transgression over much of the Slave craton. A rise greater than 300 m could have been accommodated in the modelling results by using a stronger plate. For example, a rise of 600 m could generate an acceptable fit to the observations for flexural rigidities between 4×10^{21} and 10^{23} N m, expanding our range of acceptable plate thicknesses to 8–24 km. But although a rise of 300 m is permissible, we think that anything greater is unlikely, on the basis of comparisons with the amplitude of first-order sea level changes during Phanerozoic time¹³.

The present-day flexural rigidity of cratonic North America is $\sim 7(\pm 3) \times 10^{25}$ N m, equivalent to an effective elastic plate thickness of $\sim 100 \pm 25$ km. As a rough guide, this implies a temperature of 450°C ($\pm 150^\circ\text{C}$) at a depth of ~ 100 km and

suggests that the base of the lithosphere (at a temperature of $\sim 1,300^\circ\text{C}$) lies at about 300–400 km depth. This crude estimate is consistent with the results of tomographic data, which show a high-velocity root to depths in excess of 250 km beneath cratonic North America and other cratonic areas and which suggest the presence of a thermal and chemical boundary layer (tectosphere) extending to depths of several hundred kilometres beneath cratonic regions^{14–16}. Because little relevant data is available for the Slave province for Archaean and Proterozoic time, the formation and evolution of such a boundary layer remain problematic.

Our results show that the effective elastic plate thickness of lithosphere beneath the Archaean Slave craton was smaller by about an order of magnitude in early Proterozoic time than it is today. Making the conservative assumption that the elastic strength of the continental crust is small, the 8–16 km elastic plate thickness obtained for early Proterozoic time implies a temperature of 450°C ($\pm 150^\circ\text{C}$) at a depth no more than 8–16 km below the Moho and no more than about 50 km below the surface. It is unlikely that the palaeogeothermal gradient was as high as that observed today in the Basin and Range province, because that area has a very small elastic plate thickness—no more than 5 km and perhaps as little as 1–2 km (refs 17–19). This suggests that the thermal gradient within the Slave craton at 1.9 Gyr was two to four times higher than modern thermal gradients in the same region, despite the fact that the craton became stabilized at about 2.6 Gyr (refs 20, 21). Thus the thermal boundary layer beneath the Slave craton must have thickened significantly since 1.9 Gyr, and any chemical boundary layer beneath the craton must have either developed since 1.9 Gyr, thickened significantly since 1.9 Gyr, or been at very high temperatures at 1.9 Gyr. In any case, our data preclude the possibility that a thermal boundary layer several hundred kilometres thick has existed below the Slave Craton since Archaean time. This result is significant given that data for other cratons (such as the Kaapvaal and Superior cratons) suggest, contrary to results presented here, that thick mantle roots were already in existence in Archaean time^{22,23}.

The flexural rigidity of the Slave craton at 1.9 Gyr falls just at the lower limit of the range observed in modern foreland basins (Fig. 3). We do not know if the strength of the Slave craton at 1.9 Gyr is representative of the strength for early

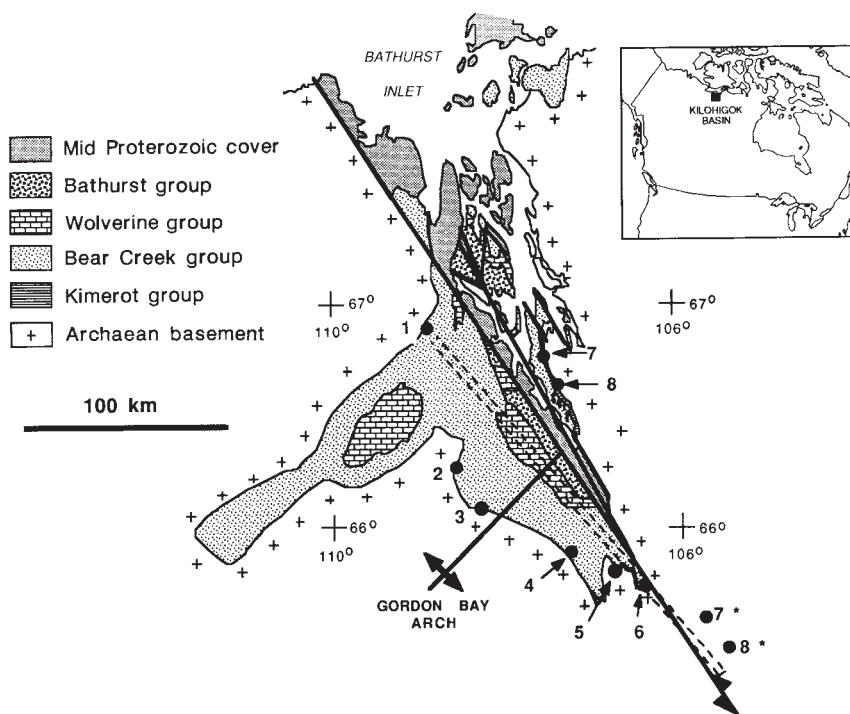


FIG. 1 Location of the Kilohigok basin, Northwest Territories, Canada. Dashed corridor represents the profile shown in Fig. 2. Stratigraphic sections were projected onto the profile along lines drawn parallel to Gordon Bay arch, which forms the outer or foreland bulge of the Kilohigok basin. 7* and 8* mark the restored positions of sections 7 and 8 after removal of 115 km of sinistral slip along the Bathurst fault^{5,8}. The Kimerot group represents the initial passive margin terrace and the Bear Creek group contains the foreland basin deposits. Thickness data for this study were derived from the lower units of the Bear Creek group (Hackett formation, Rifle formation, Beechey formation, Link formation and lower Burnside formation).

Proterozoic continental lithosphere in general. If it is a particularly weak example of early Proterozoic lithosphere, the average strength of early Proterozoic lithosphere need not have been significantly different from that today. On the other hand, we cannot rule out the possibility that the Slave craton is a particularly strong example of early Proterozoic lithosphere, in which case the average elastic plate thickness of early Proterozoic lithosphere may have been as much as an order of

magnitude less than that today, implying that continental geothermal gradients may have been a factor of two to three times greater than today. Unfortunately, the great range of flexural rigidities observed for foreland basins today and the lack of a representative sampling of ancient rigidities make it impossible to be more definitive. □

Received 12 March; accepted 20 July 1990.

1. Boak, J. & Dymek, R. F. *Earth planet. Sci. Lett.* **59**, 155–176 (1982).
2. Burke, K. & Kidd, W. S. F. *Nature* **272**, 240–241 (1978).
3. England, P. & Bickle, M. J. *J. Geol.* **92**, 353–367 (1984).
4. Watts, A. B., Bodine, J. H. & Steckler, M. S. *J. geophys. Res.* **85**, 6369–6376 (1980).
5. Grotzinger, J. P. & McCormick, D. S. in *New Perspectives in Basin Analysis* (eds Kleinspehn, K. & Paola, C.) 405–430 (Springer, New York, 1988).
6. Thompson, P. H., Culshaw, N., Buchanan, J. R. & Manojovic, P. in *Current Research, Part A Paper 86-1A*, 275–289 (Geological Survey of Canada, Ottawa, 1986).
7. van Breemen, O., Thompson, P. H., Bostock, H. H. & Loveridge, W. D. *Geol. Ass. Canada, Prog. with Abstr.* **12**, 98 (1987).
8. Tirrul, R. & Grotzinger, J. P. *Tectonics* (in the press).
9. McCormick, D. S. & Grotzinger, J. P. *Sedimentology* (submitted).
10. Grotzinger, J. P., Adams, R. D., McCormick, D. S. & Myrow, P. in *Current Research, Part C Paper 89-1C*, 107–120 (Geological Survey of Canada, Ottawa, 1989).
11. Bowring, S. A. & Grotzinger, J. P. *Geol. Ass. Canada, Prog. with Abstr.* **14**, A74 (1989).
12. Royden, L. J. *geophys. Res.* **93**, 7747–7766 (1988).
13. Kominz, M. A. in *Interregional Unconformities and Hydrocarbon Accumulation* (ed. Schlee, J. S.) 109–127 (Am. Ass. Petrol. Geol., Tulsa, 1984).
14. Grand, S. P. *J. geophys. Res.* **92**, 14065–14090 (1987).
15. Jordan, T. H. *J. Petrol. Spec. Lithosphere* **11**, 11–37 (1988).
16. Lerner-Lam, A. L. & Jordan, T. H. *Geophys. J. R. astr. Soc.* **75**, 759–797 (1987).
17. Block, L. & Royden, L. *Tectonics* (in the press).
18. McNutt, M. J. *geophys. Res.* **85**, 6377–6396 (1980).
19. Lachenbruch, A. H. & Sass, J. H. *Geol. Soc. Am. Mem.* **152**, 209–250 (1978).
20. Hoffman, P. F. in *The Geology of North America—An Overview* (eds Bally, A. W. & Palmer, A. R.) 447–512 (Geol. Soc. Am., Boulder, 1989).
21. Kusky, T. M. *Geology* **17**, 63–67 (1989).
22. Boyd, F. R., Gurney, J. J. & Richardson, S. H. *Nature* **315**, 387–389 (1985).
23. Hoffman, P. F. *Phil. Trans. R. Soc.* (in the press).
24. Bechtel, T. D., Forsyth, D. W., Sharpton, V. L. & Grieve, R. A. F. *Nature* **343**, 636–637 (1990).
25. Sheffels, B. A. M. & McNutt, M. K. *J. geophys. Res.* **91**, 6419–6431 (1986).
26. McNutt, M. K., Diament, M. & Kogan, J. G. *J. geophys. Res.* **93**, 8825–8838 (1988).
27. Royden, L. & Karner, G. D. *Nature* **309**, 142–144 (1984).
28. Karner, G. D. & Watts, A. B. *J. geophys. Res.* **88**, 10449–10477 (1983).
29. McNutt, M. K. & Kogan, M. G. in *The Composition, Structure and Dynamics of the Lithosphere–Asthenosphere System* (ed. Fuchs, K. A. F.) 309–327 (Am. geophys. Un., Washington, DC, 1987).
30. Lyon-Caen, H. & Molnar, P. *Geophys. Res. Lett.* **11**, 1251–1254 (1984).
31. Lyon-Caen, H., Molnar, P. & Suarez, G. *Earth planet. Sci. Lett.* **75**, 81–92 (1985).
32. Snyder, D. B. & Barazangi, M. *Tectonics* **5**, 361–373 (1986).
33. Kruse, S. & McNutt, M. K. *Tectonophysics* **154**, 1–17 (1988).
34. Ruppel, C. & McNutt, M. *Eos* **68**, 420 (1987).
35. Lyon-Caen, H. & Molnar, P. *J. geophys. Res.* **88**, 1251–1254 (1983).

ACKNOWLEDGEMENTS. We thank P. Hoffman for a review. Logistical and financial support were provided by the Geological Survey of Canada and the NSF.

Polydactyly in the earliest known tetrapod limbs

M. I. Coates & J. A. Clack

University Museum of Zoology, University of Cambridge,
Downing Street, Cambridge CB2 3EJ, UK

NEW specimens of the earliest known tetrapod limbs shows them to be polydactylous. The forelimb of *Acanthostega* has eight digits and the hindlimb of *Ichthyostega* has seven. Both of these come from the Upper Devonian of East Greenland, complementing the only other known Devonian tetrapod limb, that of *Tulerpeton* from Russia¹, which has six digits. The morphology of the specimens suggests that limbs with digits may have been adaptations to an aquatic rather than a terrestrial environment. The pattern of digits corresponds to a recently proposed model for limb development² in which digit number is unspecified, rather than earlier models^{3–10} which are rejected because they postulate a fixed number of elements in the ancestral limb. We challenge pentadactyly as primitive for tetrapods^{3,11}. The form of these limbs suggests early specialization in the evolution of the tetrapod limb bud.

Specimens of *Acanthostega* and *Ichthyostega* including articulated limbs, were collected by a Cambridge–Copenhagen expedition to East Greenland in 1987 (ref. 12) (Fig. 1a–f). The hindlimb of *Ichthyostega* is represented by Geological Museum

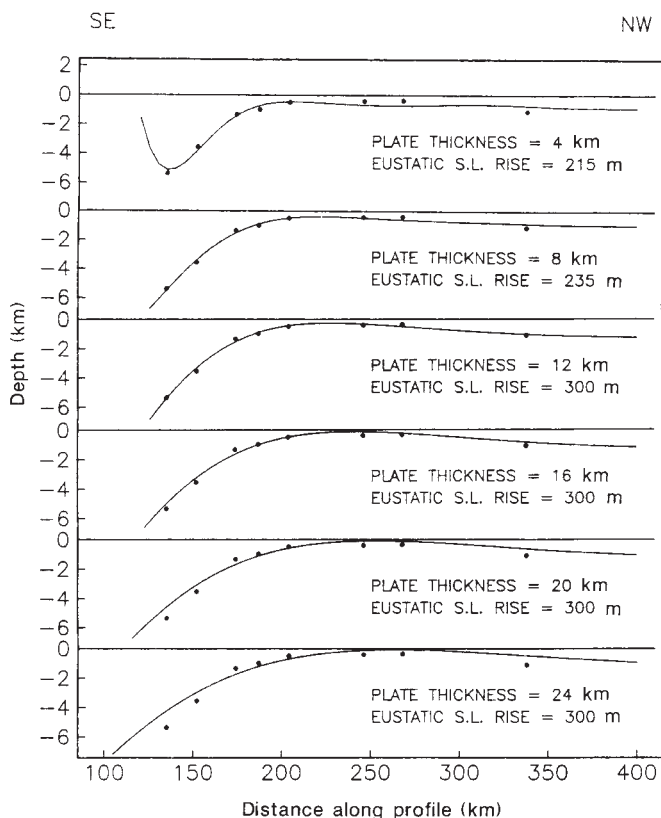


FIG. 2 Depth of the Kilohigok basin determined from stratigraphic sections at locations 1–8 in Fig. 1 and corrected for compaction (dots), and best-fitting theoretical flexural profiles for effective elastic plate thicknesses from 4 to 24 km (lines). Deposition at top and base of the section was constrained to have occurred at sea level, with a eustatic sea level rise of no more than 300 m.

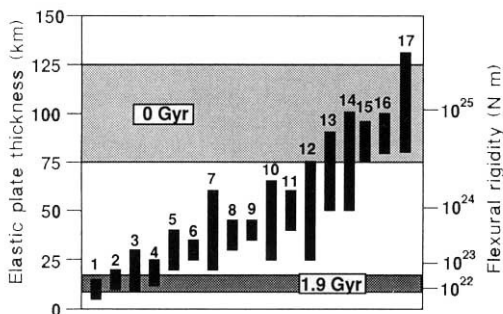


FIG. 3 Comparison of flexural rigidity (shaded) of the Slave craton at 1.9 Gyr (this paper) and for the Canadian shield at present²⁴. Vertical bars show reliable estimates of flexural rigidity for Phanerozoic foreland basins and convergent zones located within continental crust: 1, Transverse Ranges²⁵; 2, Pamirs²⁶; 3, Apennines¹²; 4, Taiwan (D. McCormick, unpublished data); 5, Carpathians²⁷; 6, Alps (ref. 28 and L. Royden, unpublished data); 7, Tien Shan²⁸; 8, East Texas (Ye Hungzhuang, unpublished data); 9, Kunlun³⁰; 10, Andes³¹; 11, Verkoyansk²⁹; 12, Zagros³²; 13, Hellenic trench (D. Dinter, unpublished data); 14, Urals^{26,33}; 15, Caucasus³⁴; 16, Himalayas³⁵; 17, Appalachians²⁸.

of Copenhagen (MGUH) specimen field number 1349, which also preserves the tail and pelvis, and the forelimb of *Acanthostega* by specimen MGUH field number 1227, which shows the dorsal surface of the left limb, with the bones well preserved and in near-perfect articulation. Two isolated humeri are also known.

The humerus of *Acanthostega* is 'L'-shaped, like those of *Ichthyostega*¹³ and other^{14,15} early but not closely related tetrapods. The proximal end bears a broad, elliptical head for articulation with the pectoral girdle. The large rectangular entepicondyle is pierced by an entepicondylar foramen; the ectepicondyle is prominent. A ridge on the dorsal surface is probably the deltoid process, and resembles that of the osteolepiform fish *Eusthenopteron*¹⁶ (Fig. 2b). Osteolepiforms are the fishes usually thought of as the sister-group of tetrapods (see, for example, ref. 17). A short canal, the ectepicondylar foramen¹⁶, corresponds to 'canal a' of the ichthyostegid humerus¹³. Distally, the humerus bears terminal facets for the radius and ulna. The radius is spatulate, broadened and flattened distally, and almost twice the length of the ulna. The ulna is short and broad, without any form of olecranon process. The carpus is poorly preserved, except for a well-ossified, sub-cylindrical bone, probably the intermedium. Eight digits lie in a broad arch, and from anterior to posterior show a phalangeal formula of 3, 3, 3, 4, 4, 5, 5, 4 (assuming minimal restoration). There are no complete hindlimbs so far known for *Acanthostega*, but the femur, tibia and fibula are similar to those of *Ichthyostega*.

The proximal region of the hindlimb of *Ichthyostega* (Fig. 1d, e, f), corresponds closely with the published description¹³, but the tarsus and digits differ. In addition to the elements described, there is a more extensive series of distal tarsals, three small digits replace digit I plus a tibiale and prehallux. Digits conventionally numbered¹³ as III and IV each contain an additional phalanx. Thus the limb bears seven digits with the phalangeal formula 3, 4, 2, 3, 4, 4, 3. The original material in Stockholm shows the same phenomenon.

The radius and ulna of *Acanthostega* (Fig. 1a, b, c) are of remarkably fish-like proportions (Fig. 2a, b)^{16,18}. No other tetrapod is known to show such fish-like lower limb bones, and we interpret this as a primitive character. This, and the evidence of polydactyly in all three known Devonian tetrapods, persuades us that pentadactyly is not primitive for tetrapods. Evidence for pentadactyly in early tetrapods is scarce. Only four genera from Lower Carboniferous (Mississippian) deposits show articulated pentadactyl hindlimbs, and only two show articulated forelimbs, bearing respectively four and five digits^{15,19-21}.

A recent analysis of tetrapod relationships has suggested that tetrapods divided early into two clades²² with *Ichthyostega* at the base of one of these. Other members of both these clades have pentadactyl hindlimbs, whereas the clade including *Ichthyostega*, where known, has only four digits on the manus (the manus of *Ichthyostega* remains unknown). Under this scheme pentadactyly was presumably convergently derived in the two clades, and a five-digit manus may never have been present in one lineage.

The morphology of the limbs of *Acanthostega* and *Ichthyostega* suggest an aquatic mode of life, compatible with a recent reassessment of functional aspects of the fish-tetrapod transition²⁴. The dorsoventrally compressed lower leg bones of *Ichthyostega* strongly resemble those of a cetacean pectoral flipper. A peculiar, poorly ossified mass lies anteriorly adjacent to the digits, and appears to be reinforcement for the leading edge of this paddle-like limb (Fig. 1d, e, f). The tarsus of *Ichthyostega* may have been inflexible²³, consistent with the previous suggestion. But the large forelimb of *Ichthyostega* is, by contrast, modified into an apparently load-bearing structure¹³.

In *Acanthostega* the terminal radial and ulnar condyles, and the lack of an olecranon process on the ulna, suggest that the forelimb could never have flexed from the elbow to lie in a fully load-bearing posture, and that it was probably held horizontally.

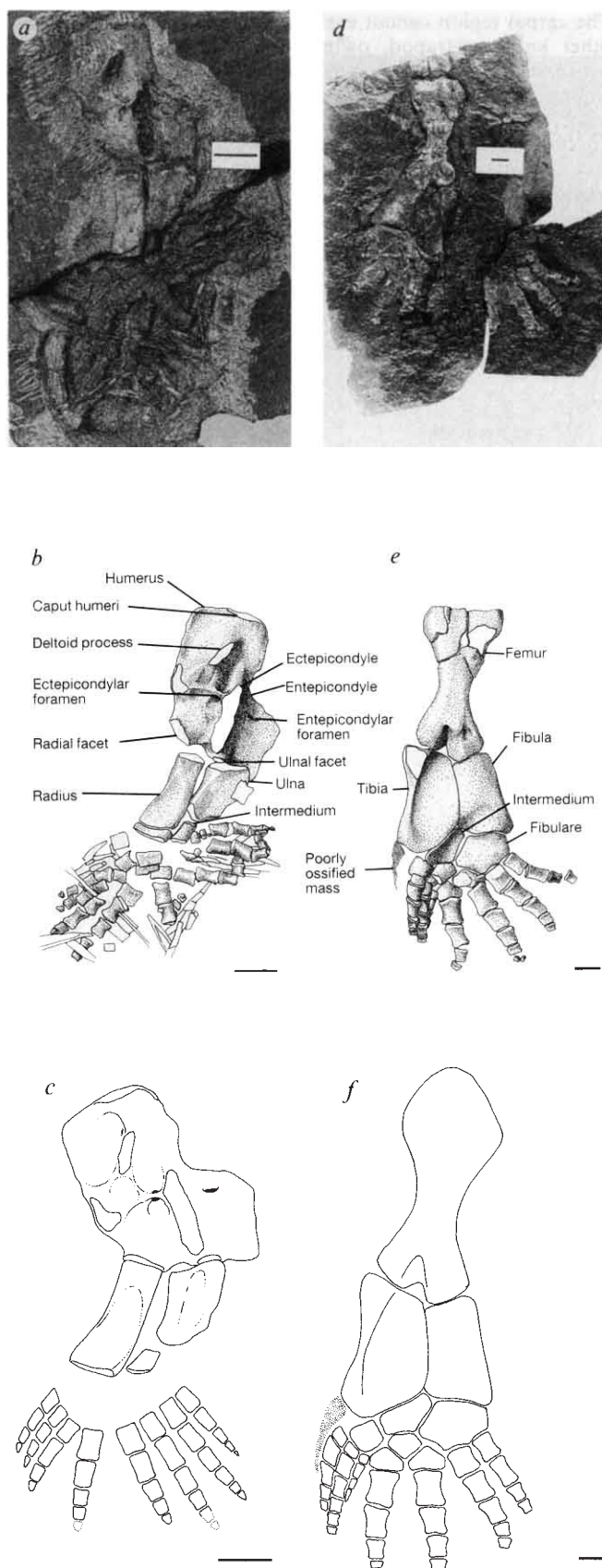


FIG. 1 Limb skeletons, a, *Acanthostega* forelimb (specimen 1227). b, Drawing of forelimb. c, Restoration of forelimb. d, *Ichthyostega* hindlimb (specimen 1349, part and counterpart halves). e, Composite drawing of hindlimb based on data from part and counterpart halves. f, Restoration of hindlimb. All limbs in dorsal view; anterior edges to left. All scale bars, 10 mm.

The carpal region cannot easily be compared with that of any other known tetrapod, owing to the discrepancy in length between the radius and ulna, and is functionally difficult to explain. But the development of digits while retaining an aquatic habit has long been known to occur²⁴. Limb-like fins with digit-like rays are used for grasping, stalking, or as supplementary sensory organs (for example, the pseudo-digitated fins of triglids (gurnards), *Lophius* (anglers), *Histrio* (sargassum frog-fishes), and ogocephalids (bat-fishes)). Digits in early tetrapods may have functioned similarly.

Almost all attempts to find homologies between tetrapod limb bones and the fin-bones of lobe-finned fishes³⁻¹⁰ (Fig. 2*b, c*) have been based on the assumption that all tetrapod limbs are derived from a fundamental and fixed number of 'canonical'^{3,8}

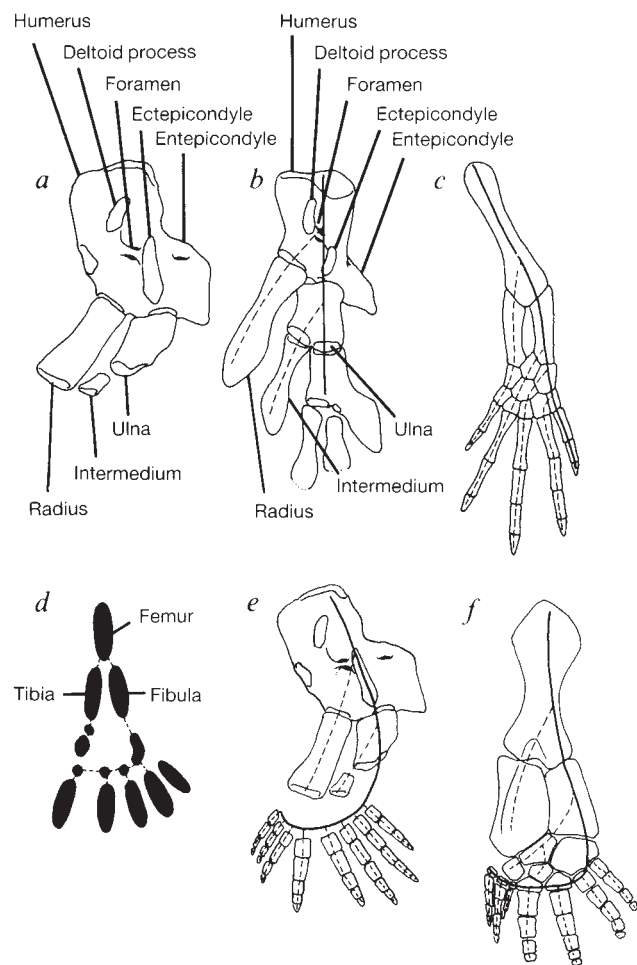


FIG. 2 *a*, *Acanthostega* pectoral limb skeleton. Note similarities with structures shown in *b*, *Eusthenopteron* pectoral fin skeleton (after Andrews and Westoll¹⁶, and University Museum of Zoology, Cambridge, specimen GN190), metapterygial axis of fin indicated by heavy line; radials (pre-axial) indicated by dashed lines. Note that metapterygial axis cannot be plotted with any certainty into either of the most distal bones; therefore post-axial radial potentially present. *c*, Gegenbaur's, 1876 (ref. 4), plan of an idealized limb, the earliest of such representations of hypotheses of limb ontogeny and phylogeny. Heavy line, metapterygial axis; dashed line, radial axes. This assumes a fixed primitive pattern of tetrapod limb bones and homology of fin radials to digits combined with wrist/ankle bones (see also refs 5-10; all lie within a recapitulationist (von Baerian or Haeckelian) paradigm). *d*, Schematic diagram of hindlimb development in *Xenopus laevis* (after Shubin and Alberch² Fig. 11). Dashed lines indicate embryonic (cellular) connections between pre-cartilage blastemata in the developing limb-bud. Note form of digital arch and conventional numbering of digits. *e, f*, Acanthostegid and ichthyostegid limbs interpreted after the inferred fin axes of Shubin and Alberch² (adapted from ref. 2, Fig. 20*a*). All limbs/fins in dorsal view; anterior edges to left.

elements and in which the main (metapterygial) axis runs through or between digits. But no known fin could provide sufficient elements for the distal parts of the tetrapod limb¹⁸, none of these 'ancestral patterns' can be found in the ontogeny of extant tetrapod limbs²⁵⁻²⁸, and none of these patterns can account for the limb morphology of *Acanthostega* and *Ichthyostega*. An alternative view of limb development²⁹ rejects the concept of morphological homology in favour of recognition of the unity of underlying morphogenetic rules. Significantly, a recent analysis of limb development has found it subject to three morphogenetic rules: segmentation, bifurcation, and focal condensation², and suggests that the main (metapterygial) axis runs round the wrist or ankle (Fig. 2*d*). Thus the most distal structures developmentally, lie at the anterior edge of the limb. Only the radius/tibia, intermedium and some of the wrist and ankle bones are pre-axial, while the digits are all post-axial.

Loss or failure to develop digits often occurs at the distal end of the digital arch. Additional digits in modern polydactylous forms often arise distally to digit 1, at the anterior edge of the limb^{30,31}. Thus in *Acanthostega* and *Ichthyostega*, the smallest and shortest digits in the array are closest to the anterior edge. This pattern could have been predicted and is elegantly explained by the new morphogenetic model² (Fig. 2*e, f*). Pentadactyly would therefore be derived by loss of the most terminal digits.

This approach cannot resolve the debate over whether lungfishes or osteolepiforms are the sister-group of tetrapods^{24,32}. Although the tetrapod limb could be conceived of as an anteriorly curled fin of either of these fish groups, these 'conserved generative rules'²⁵ seem to be common to all sarcopterygians, the group containing lungfishes, osteolepiforms and tetrapods. Synapomorphic morphological similarities of the humeri (Fig. 2*a, b*) and the limb joints remain more precise indicators of the closer relationships of the osteolepiforms with tetrapods¹⁷.

The proposal of this underlying pattern for limb development², and its independent corroboration by fossil data (on the basis of observable processes, the acanthostegid and ichthyostegid limb morphologies are irreducible to any of the other competing schemes of limb morphogenesis) provides developmental biology with an evolutionary perspective. It has implications for the conservation of the role of the zone of polarizing activity^{27,30,33,34} in limb pattern formation; primitive polydactyly suggests changes in limb-bud topography^{28,30,33,35,36}. Therefore the establishment of pentadactyly may result from both developmental and functional constraints. The hindlimb of *Ichthyostega* demonstrates what has been considered a 'forbidden morphology' in developmental terms^{29,37}. Vertebrate palaeontology thus provides a unique contribution to the study of limb development, and illustrates how the study of Devonian tetrapods is relevant to understanding the structure of modern animals. □

Received 8 May; accepted 25 June 1990.

- Lebedev, O. A. *Dokl. Akad. Nauk SSSR* **278**, 1407-1473 (1984).
- Shubin, N. H. & Alberch, P. *Evol. Biol.* **20**, 319-387 (1986).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. 2 (Academic, London, 1980).
- Gegenbaur, C. *Morph. Jb.* **2**, 396-420 (1876).
- Watson, D. M. S. *Anat. Anz.* **44**, 24-27 (1913).
- Gregory, W. K., Miner, R. W. & Noble, G. K. *Bull. Am. Mus. nat. Hist.* **48**, 279-288 (1923).
- Gregory, W. K. & Raven, H. C. *Ann. N.Y. Acad. Sci.* **42**, 273-360 (1941).
- Holmgren, N. *Acta zool.* **14**, 185-295 (1933).
- Westoll, T. S. *Proc. R. Soc. B* **131**, 373-393 (1943).
- Steiner, H. *Rev. Suisse. Zool.* **41**, 383-396 (1934).
- Goodrich, E. S. *Studies on the Structure and Development of Vertebrates* (Macmillan, London, 1930).
- Panchen, A. L. *Nature* **333**, 704 (1988).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. 1 (Academic, London, 1980).
- Smithson, T. R. *Zool. J. Linn. Soc.* **85**, 317-410 (1985).
- Holmes, R. *Phil. Trans. R. Soc. B* **306**, 413-527 (1984).
- Andrews, S. M. & Westoll, T. S. *Trans. R. Soc. Edinb.* **68**, 207-329 (1970).
- Panchen, A. L. & Smithson, T. R. *Biol. Rev.* **62**, 341-438 (1987).
- Rackoff, J. S. in *The Terrestrial Environment and the Origin of Land Vertebrates* (ed. Panchen, A. L.) 255-292 (Academic, London, 1980).
- Wood, S. P., Panchen, A. L. & Smithson, T. R. *Nature* **314**, 355-356 (1985).

20. Smithson, T. R. *Nature* **342**, 676–678 (1989).
21. Godfrey, S. J. *Phil. Trans. R. Soc. B* **323**, 75–133 (1989).
22. Panchen, A. L. & Smithson, T. R. in *The Phylogeny and Classification of the Tetrapods. Volume 1: Amphibians, Reptiles, Birds* (ed. Benton, M. J.) 1–32 (Clarendon, Oxford, 1988).
23. Bjerring, H. C. in *Evolutionary Biology of Primitive Fishes* (eds Foreman, R. E. et al.) (Plenum, New York, 1985).
24. Edwards, J. L. *Am. Zool.* **29**, 235–254 (1989).
25. Müller, G. B. & Alberch, P. *J. Morph.* **203**, 151–164 (1990).
26. Hinchliffe, J. R. & Griffiths, P. J. in *Development and Evolution* (eds Goodwin, B. C. et al.) 99–121 (Cambridge University Press, Cambridge, 1983).
27. Hinchliffe, J. R. in *Fortschritte der Zoologie Vol. 35* (eds Spiechtna, H. & Hilgers, C. H.) 226–229 (Gustav Fischer Verlag, Stuttgart, 1989).
28. Oster, G. F. et al. *Evolution* **42**, 862–884 (1988).
29. Goodwin, B. C. & Trainor, L. E. H. in *Development and Evolution* (eds Goodwin, B. C. et al.) 75–98 (Cambridge University Press, Cambridge, 1983).
30. Hinchliffe, J. R. in *Complex Organismal Functions: Integration and Evolution in Vertebrates* (eds Wake, D. B. & Roth, G.) 171–189 (Wiley, Chichester, 1989).
31. Danforth, C. H. *Am. J. Anat.* **80**, 143–171 (1947).
32. Rosen, D. E. et al. *Bull. Am. Mus. nat. Hist.* **67**, 159–276 (1981).
33. Brickell, P. M. & Tickle, C. *BioEssays* **11**, 145–149 (1989).
34. Wolpert, L. in *Development and Evolution* (eds Goodwin, B. C. et al.) 47–51 (Cambridge University Press, Cambridge, 1983).
35. Saunders, J. W. in *Vertebrate Limb and Somite Morphogenesis* (eds Ede, D. A. et al.) 1–24 (Cambridge University Press, Cambridge, 1977).
36. Hinchliffe, J. R. & Johnson, D. R. *The Development of the Vertebrate Limb* (Clarendon, Oxford, 1980).
37. Holder, N. J. *theor. Biol.* **104**, 451–471 (1983).

ACKNOWLEDGEMENTS. We are grateful to Erik Jarvik for allowing us to examine the original ichthyostegid material, and providing facilities for its study at the Paleozoology Section, Swedish Museum of Natural History, Stockholm. We thank C. M. Bate, A. E. Friday, A. Martinez Arias, T. R. Smithson and T. G. Benton for helpful criticism of earlier drafts of the manuscript, and N. Fraser for discussions and technical assistance. We thank the Geological Survey of Greenland, Denmark for practical support in the field, and the Carlsberg Foundation (Copenhagen), the Natural Environment Research Council, and the University Museum of Zoology, Cambridge for financial support.

Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex

A. Artola, S. Bröcher & W. Singer

Max-Planck-Institute for Brain Research, Deutschordenstrasse 46,
D-6000 Frankfurt am Main 71, FRG

IN the hippocampus and neocortex, high-frequency (tetanic) stimulation of an afferent pathway leads to long-term potentiation (LTP) of synaptic transmission^{1–5}. In the hippocampus it has recently been shown that long-term depression (LTD) of excitatory transmission can also be induced by certain combinations of synaptic activation^{6,7}. In most hippocampal⁸ and all neocortical pathways^{4,9–11} studied so far, the induction of LTP requires the activation of *N*-methyl-D-aspartate (NMDA) receptor-gated conductances. Here we report that LTD can occur in neurons of slices of the rat visual cortex and that the same tetanic stimulation can induce either LTP or LTD depending on the level of depolarization of the postsynaptic neuron. By applying intracellular current injections or pharmacological disinhibition to modify the depolarizing response of the postsynaptic neuron to tetanic stimulation, we show that the mechanisms of induction of LTD and LTP are both postsynaptic. LTD is obtained if postsynaptic depolarization exceeds a critical level but remains below a threshold related to NMDA receptor-gated conductances, whereas LTP is induced if this second threshold is reached.

In slices of adult rat visual cortex, intracellular recordings were obtained from regular spiking cells in layers II and III to assess the effect of tetanic (50 Hz) stimulation of underlying white matter on postsynaptic potentials (p.s.ps) evoked by stimulation (0.03 Hz) of white matter or of cortical layer II 1.5–2 mm laterally to the recorded neuron (intracortical).

In a first series of experiments we investigated post-tetanic modifications of the tetanized pathway (white matter) in slices treated with various concentrations of bicuculline (0.1–0.3 μ M). In normal medium ($n = 12$) there were no modifications of p.s.ps

(Fig. 1a, d). In 0.1–0.2 μ M bicuculline ($n = 7$), tetanic stimulation induced LTD (Fig. 1b, d) and in 0.3 μ M bicuculline it led to LTP ($n = 20$). However, in eight cells only a late component of the p.s.p. measured at 22 ms post-stimulus had increased without any modification of the early p.s.p. component (Fig. 1d and legend to Fig. 1; see also ref. 11). The remaining 12 cells also showed a potentiation of the early p.s.p. (Fig. 1c, d). All post-tetanic modifications of the early p.s.p. amplitude were associated with a corresponding change of the initial slope of the p.s.p. (Fig. 1b(3), c(3)). These results suggest that slight disinhibition raises the probability that tetanic stimulation induces LTD, whereas stronger disinhibition favours the occurrence of LTP (Fig. 1e).

Disinhibition enhances the depolarizing response by two mechanisms. First, it reduces Cl^- conductances in the recorded neuron and second, by increasing activity in polysynaptic pathways¹² it increases the excitatory input. By analogy with the mechanism for LTP induction¹³, this suggests that the mechanism underlying LTD induction also involves a voltage-dependent threshold in the postsynaptic neuron which is, however, lower than that for LTP induction. As the latter involves NMDA receptor-gated conductances^{4,9–11}, we hypothesized that the process leading to LTD is related to excitatory mechanisms other than the NMDA receptors.

To test this hypothesis, slices were treated with 0.3 μ M bicuculline to raise excitability and at the same time NMDA receptors were blocked by 25 μ M 2-amino-5-phosphonovaleric acid (APV). In 10 out of 11 cells, tetanic stimulation induced LTD ($79.5 \pm 6.4\%$) which persisted after wash-out of APV (Fig. 2a, b) and was associated with a decrease of the initial slope of the p.s.p. (Fig. 2b, inset). Compared with the corresponding values in cells kept in normal medium, this depression was significant ($P < 0.05$). The remaining cell exhibited LTP but this developed more slowly than in cells treated only with 0.3 μ M bicuculline. LTD of similar magnitude ($82.0 \pm 5.7\%$, $n = 3$) was obtained when APV concentration was raised to 100 μ M. After APV had been washed out, a second tetanic stimulus (more than 1 h after the first) induced LTP ($n = 8$; data not shown). Tetanic stimuli had no effect on responses recorded from cells treated only with APV ($n = 5$). Induction of LTD was specific to the tetanized (white matter) pathway. Responses evoked from intracortical stimulation were not modified in seven out of eight cells in which LTD had developed in the tetanized pathway (Fig. 2c–e). In the remaining cell, both responses were similarly depressed. We assume that here the same intracortical pathways were activated by the two stimulating electrodes. These results demonstrate that LTD does not involve NMDA receptors, is selective for the tetanized pathway and can be reversed by LTP-inducing stimuli.

In a second series of experiments we increased postsynaptic activation by raising the intensity of the tetanic stimulus (four times the test intensity) and obtained LTD also in slices kept in normal medium ($67.2 \pm 2.6\%$, $n = 13$; Fig. 3a, e). As before, this depression was associated with a decrease of the initial slope of the p.s.p. (Fig. 3a(3)), remained inducible in the presence of 25 μ M APV and was confined to the tetanized afferents (Fig. 3e). As a direct test of the hypothesis that LTD and LTP have different voltage-dependent thresholds, prolonged currents (1 min, 0.4–0.6 nA) were injected throughout tetanic stimulation to produce hyper- and depolarizing shifts of V_m of -40 mV, $+10$ mV and $+20$ mV, respectively. Hyperpolarizing ($n = 6$; Fig. 3b, e) but also slight ($+10$ mV) depolarizing currents ($n = 2$; data not shown) consistently prevented the occurrence of LTD, but application of stronger depolarizing currents ($+20$ mV) led to LTP in nine out of ten cells ($127.2 \pm 4.2\%$, $n = 9$; Fig. 3c, e). This potentiation was prevented when 25 μ M APV was present during the tetanus ($n = 6$, data not shown). Both LTD and LTP could be reversed (Fig. 3d). These results indicate that the changes leading to LTD and LTP had actually occurred at synapses located on the recorded neuron and were selective for

the tetanized afferents. Moreover they provide direct proof that the processes leading to LTD and LTP both depend on postsynaptic depolarization and have different thresholds (Fig. 3f).

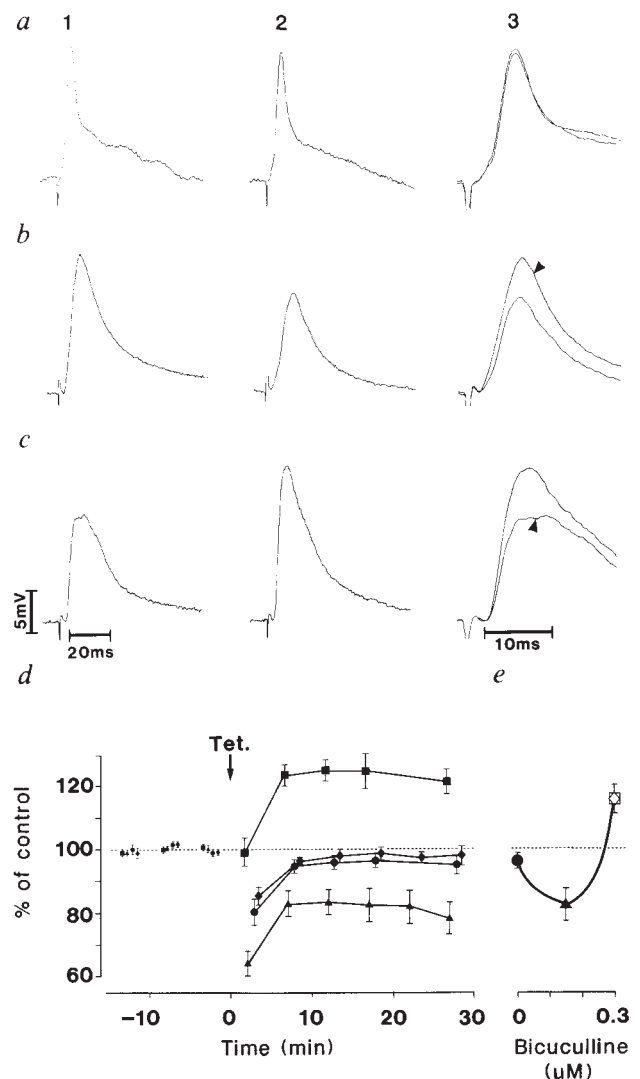
In conclusion, and as illustrated schematically in Fig. 3f, a constant synaptic input can lead either to no change, to LTD or to LTP of the activated synapses depending on the level of depolarization of the postsynaptic neuron. This predicts that there are two ranges of postsynaptic activation where no change is expected; one is below LTD threshold and the other somewhere between LTD and LTP threshold. We suggest that the fraction of neurons treated with 0.3 μM bicuculline which showed only potentiation of a late but not of the early p.s.p. (Fig. 1d) and the fraction of neurons that were only slightly depolarized (+10 mV) during the tetanus were in this second range.

The mechanism proposed here accounts for recent reports of homosynaptic LTD in hippocampus, where established LTP can be reversed if the potentiated pathway is stimulated at 1 Hz⁷, and LTD can be induced *de novo* when synaptic activation at 5 Hz occurs while the postsynaptic neuron is slightly hyperpolarized⁶. We propose that in both cases LTD was obtained because activation was above LTD and below LTP threshold, in the first case because of low temporal summation and in the second

FIG. 2 Effect of tetanic stimulation on synaptic responses in slices treated with 0.3 μM bicuculline and 25 μM APV. *a*, Post-tetanic changes of p.s.p. amplitudes (white matter stimulation) measured at +11 ms post-stimulus. Tetanic stimulation was applied to white matter 90 min after bath application of 0.3 μM bicuculline and 20 min after 25 μM APV. Bicuculline was present throughout the experiment; APV was discontinued 20 min after tetanic stimulation. Values are expressed as per cent of response amplitude measured before the beginning of APV perfusion. *b*, Postsynaptic potentials at times 1–4 indicated in *a*. The inset represents the superimposed responses 2, 3 and 4 at expanded time and amplitude scales. Note that washout of APV does not modify the slope of the response. $V_m = -77$ mV. *c–d*, Postsynaptic potentials elicited by white matter (tetanized pathway) (*c*) and intracortical stimulation (*d*) before (i) and 20 min after (ii) the tetanus. Responses i and ii are superimposed at expanded timescale in iii. Arrows indicate the pretetanic response. $V_m = -70$ mV. Each trace in *b–d* is the average of five successive responses (0.03 Hz). *e*, Post-tetanic changes of p.s.p. amplitude of responses to white matter (●) and intracortical (○) stimulation. Values represent averages from seven cells and are expressed as per cent pretetanic controls. Vertical bars, $\pm$ s.e.m. The amplitude change in the tetanized pathway ($75.4 \pm 7.3\%$) differs significantly ($P < 0.05$) from that in the non-tetanized pathway (intracortical stimulation; $92.7 \pm 2.4\%$).

FIG. 1 Effect of tetanic stimulation (white matter) in slices treated with various concentrations of bicuculline. *a–c*, Synaptic responses to white matter stimulation before (1) and 20 min after (2) tetanic stimulation in normal medium (*a*), in 0.1 μM (*b*) and 0.3 μM (*c*) bicuculline. Responses 1 and 2 are superimposed in 3 at expanded timescale to show the modifications of the initial slope. Arrows indicate responses before tetanic stimulation. Each record in *a–c* is the average of five successive responses (0.03 Hz). $V_m = -70$ mV in *a*, -79 mV in *b* and -71 mV in *c*. *d*, Time course of the post-tetanic changes in early p.s.p. amplitude of cells in normal medium ($n=12$; ●), of cells in 0.1–0.2 μM bicuculline ($n=7$; ▲) and of cells in 0.3 μM bicuculline showing LTP of the late p.s.p. component only ($n=8$; ◆) and of both p.s.p. components ($n=12$; ■). As described previously¹¹, cells showing LTP of both p.s.p. components differed from those exhibiting LTP of only the late p.s.p. component in that bicuculline treatment was more effective in the former than in the latter. Values in *d* are averages of early p.s.p. amplitudes for each group expressed as per cent pretetanic controls (see methods below). Vertical bars show $\pm$ s.e.m.. *e*, Schematic representation of the dependence of post-tetanic p.s.p. modifications (ordinate, measured 20 min after tetanus) on bicucullin concentration (abscissa). For this plot all cells receiving the same pharmacological treatment have been added together including those that failed to exhibit LTP of the early p.s.p. in 0.3 μM bicuculline. Values of cells treated with 0.1–0.2 μM bicuculline ($81.7 \pm 4.5\%$; $n=7$; ▲) and with 0.3 μM bicuculline ($115.5 \pm 4.6\%$; $n=20$; ◆), respectively are significantly ($P < 0.01$) different from those of cells in normal medium ($97.5 \pm 1.7\%$, $n=12$, ●). Resting membrane potential, input resistance and stimulation intensity did not differ between groups.

METHODS. Neocortical slices prepared from visual cortex of adult rats by standard methods were incubated in a submerged slice chamber at 29–30 °C. The superfusion medium contained NaCl 124 mM, KCl 5 mM, NaHPO₄ 1.25 mM, MgSO₄ 2 mM, CaCl₂ 2 mM, NaHCO₃ 26 mM, D-glucose 10 mM, and H₂O₂ 0.002% bubbled with 95% O₂: 5% CO₂. Intracellular recordings were obtained with electrodes filled with 3 M potassium acetate (80–120 M Ω). Stimuli (pulse duration 200 μs) were applied with a bipolar tungsten electrode. Tetanic stimuli were five, 2-s stimulus trains (50 Hz) at 10-s intervals. During tetanic stimulation, intensity was raised to suprathreshold levels, twice the test intensity in experiments documented in Figs 1 and 2 and four times the test intensity in experiments illustrated in Fig. 3. Bicuculline-containing medium was applied for more than 1 h before tetanic stimulation. To assess post-tetanic changes in the early part of the test-response (early p.s.p.) which contains the monosynaptic e.p.s.p., we measured the rising slope and the amplitude at peak or at 11 ms post-stimulus when the peak was delayed (which occurred with 0.3 μM of bicuculline). In addition we measured p.s.p. amplitude at 22 ms post-stimulus to account for cases where tetanic stimulation affected only a late component of the p.s.p.¹¹. Quantitative data on post-tetanic p.s.p. modifications are derived exclusively from measurements of the early p.s.p. component and are expressed as percentage of controls, the latter representing the mean of responses recorded during the stable period 15–30 min before tetanic stimulation. Unless stated otherwise, values given in the text are the mean $\pm$ s.e.m. of



changes in the respective cell populations measured 20 min after tetanus. The Student *t* test or the Mann-Whitney *U* test were applied to compare the means.

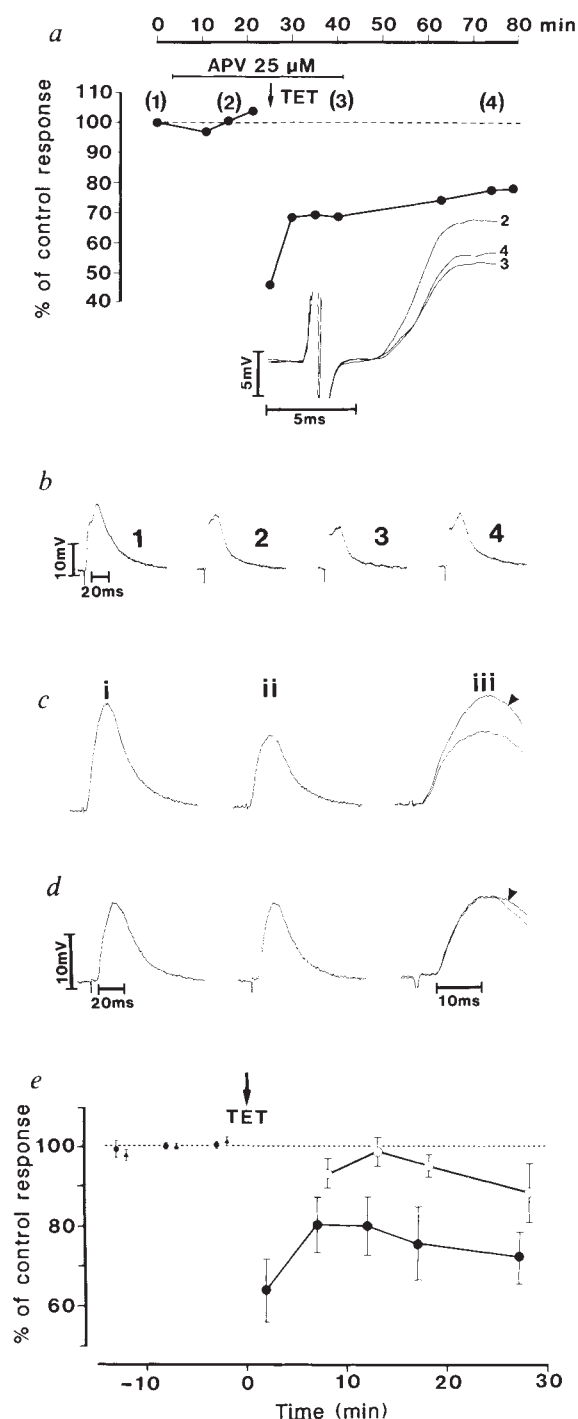
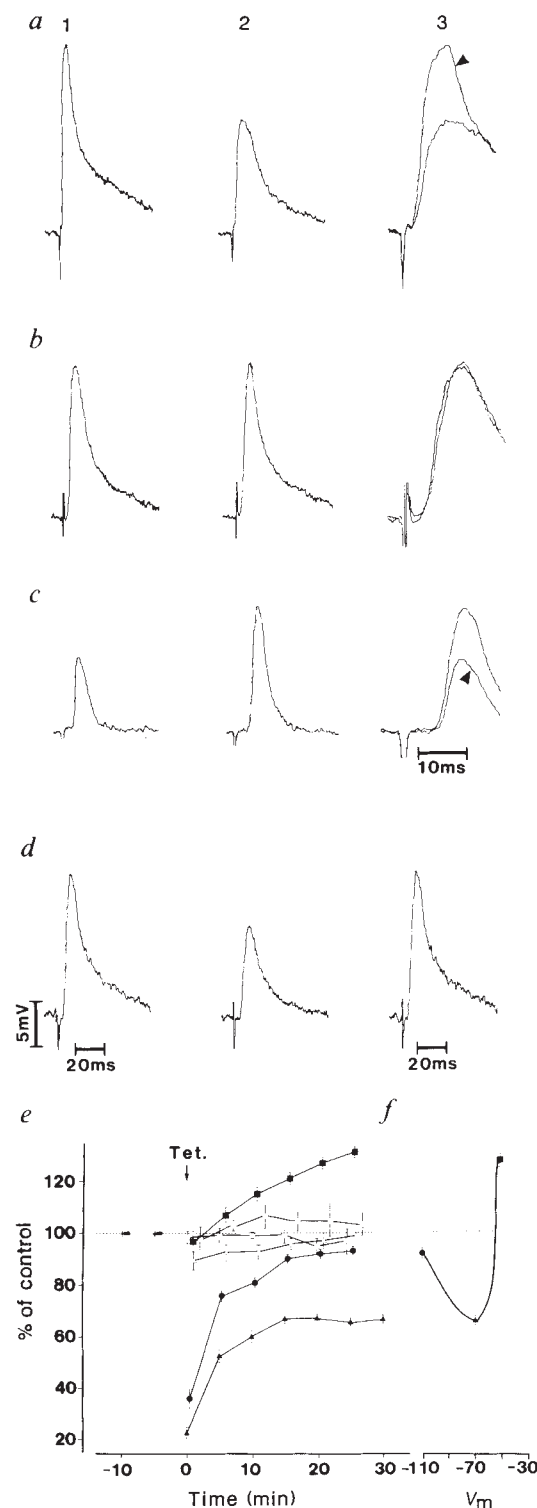


FIG. 3 Effect of depolarizing and hyperpolarizing current pulses on induction of LTD and LTP. Intensity of the tetanic stimulus was four times the test intensity. *a-c*, Averaged ($n=5$, 0.03 Hz) responses to white matter stimulation recorded before (1) and 20 min after (2) the tetanus from a cell receiving no current injection during the tetanus (*a*), in a cell hyperpolarized by -40 mV below V_{mr} (*b*) and in a cell depolarized by $+20$ mV above V_{mr} (*c*). Responses 1 and 2 are superimposed at expanded timescale in 3. Arrows indicate the pretetanic response. *d*, Induction and reversal of LTD. A tetanus was applied without current injection and caused LTD (compare control, 1, with the post-tetanic response, 2). A second tetanus was applied in conjunction with a depolarizing current pulse and this caused LTP of the previously depressed response resulting in a resetting of LTD (3). $V_m = -71$ mV in *a*; -69 mV in *b*; -71 mV in *c* and -73 mV in *d*. *e*, Time course of post-tetanic changes of p.s.p. amplitude in cells having received no current injection during tetanus ($\blacktriangle$, $n=13$), and in cells conditioned with hyperpolarizing ($\bullet$, $n=6$) and



depolarizing ($\blacksquare$, $+20$ mV, $n=9$) currents, respectively. Conventions are as in Fig. 1*d*. Open symbols, amplitudes of the response to intracortical stimulation for the corresponding groups. *f*, Schematic representation of the dependence of post-tetanic ($+20$ min after tetanus) p.s.p. modifications on V_m levels (abscissa) induced by current injection. Note that these V_m levels change during the tetanus because of additional action of synaptic currents and modifications of membrane conductance. The values of the cell groups showing LTD ($\blacktriangle$) and LTP ($\blacksquare$) differed significantly ($P<0.001$) from the corresponding values of cells which had been hyperpolarized ($\bullet$, $92.8 \pm 2.8\%$) or slightly ($+10$ mV) depolarized ($101.5 \pm 0.2\%$, $n=2$, not shown) during tetanus. They are also significantly ($P<0.001$) different from the corresponding values of p.s.p. amplitude in the non-tetanzed pathway (intracortical stimulation). They are $97.2 \pm 2.8\%$ (Δ), $94.9 \pm 2.1\%$ ($\circ$) and $105 \pm 6.5\%$ ($\square$), respectively.

because of hyperpolarization. Our results may also have some bearing on developmental plasticity. They can explain why in monocularly deprived kittens, the afferent inputs from the open, more active eye are repressed if cortical neurons are prevented from responding by application of either the GABA-A receptor agonist muscimol¹⁵ or high doses of APV¹⁶. This could occur if dendritic responses reach the threshold for LTD but the muscimol or APV treatment prevented the activation of the NMDA receptor-mediated mechanisms. Furthermore, our results provide experimental support for the proposal that the sign of use-dependent synaptic modifications in the kitten visual cortex switches from negative to positive as a function of the activation level^{17,18}. For *in vivo* conditions, we predict that the sign of use-dependent synaptic modifications depends on the balance between contingent excitatory and inhibitory influences on a neuron. Thus, both LTD and LTP have an associative function, but LTD works by a process inverse to the normal hebbian rule involved in LTP. This inverse-hebbian mechanism may be involved in learning processes which lead to weakening of stimulus-response relations such as habituation, reversal learning and extinction. □

Received 28 November 1989; accepted 28 June 1990.

1. Bliss, T. V. P. & Lomo, T. *J. Physiol., Lond.* **232**, 331–356 (1973).
2. Komatsu, Y., Toyama, K., Maeda, J. & Sakaguchi, H. *Neurosci. Lett.* **26**, 269–274 (1981).
3. Lee, K. S. *Brain Res.* **239**, 617–623 (1982).
4. Artola, A. & Singer, W. *Nature* **330**, 649–652 (1987).
5. Perkins, A. T. & Teyler, T. J. *Brain Res.* **439**, 222–229 (1988).
6. Stanton, P. K. & Sejnowski, T. J. *Nature* **339**, 215–218 (1989).
7. Staubli, U. & Lynch, G. *Neurosci. Abstr.* **225**, 16 (1988).
8. Collingridge, G. L., Kehl, S. J. & McLennan, H. *J. Physiol., Lond.* **334**, 33–46 (1983).
9. Sutor, B. & Häblitz, J. *J. Neurosci. Lett.* **97**, 111–117 (1989).
10. Kimura, F., Nishigori, A., Shirokawa, T. & Tsumoto, T. *J. Physiol., Lond.* **414**, 125–144 (1989).
11. Artola, A. & Singer, W. *Eur. J. Neurosci.* **2**, 254–269 (1990).
12. Chagnac-Amitai, Y. & Connors, B. W. *J. Neurophysiol.* **61**, 747–758 (1989).
13. Wigström, H. & Gustafsson, B. *Nature* **301**, 603–604 (1983).
14. Larson, J. & Lynch, G. *Science* **232**, 985–988 (1986).
15. Reiter, H. O. & Stryker, M. P. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3623–3627 (1988).
16. Bear, M. F., Kleinschmidt, A., Gu, Q. & Singer, W. *J. Neurosci.* **10**, 909–925 (1990).
17. Bienenstock, E. L., Cooper, L. N. & Munro, P. W. *J. Neurosci.* **2**, 32–48 (1982).
18. Bear, M. F., Cooper, L. N. & Ebner, F. F. *Science* **237**, 42–48 (1987).

Molecular cloning and expression of the gene for a human D₁ dopamine receptor

Allen Dearry^{*†}, Jay A. Gingrich^{*}, Pierre Falardeau^{*}, Robert T. Freneau Jr[‡], Michael D. Bates^{*} & Marc G. Caron^{*§||}

Departments of ^{*}Cell Biology, [†]Ophthalmology, [‡]Neurobiology, and ^{||}Medicine, and [§]Howard Hughes Medical Institute Laboratories, Duke University Medical Center, Durham, North Carolina 27710, USA

THE diverse physiological actions of dopamine are mediated by its interaction with two basic types of G protein-coupled receptor, D₁ and D₂, which stimulate and inhibit, respectively, the enzyme adenylyl cyclase¹. Alterations in the number or activity of these receptors may be a contributory factor in diseases such as Parkinson's disease and schizophrenia. Here we describe the isolation and characterization of the gene encoding a human D₁ dopamine receptor. The coding region of this gene is intronless, unlike the gene encoding the D₂ dopamine receptor. The D₁ receptor gene encodes a protein of 446 amino acids having a predicted relative molecular mass of 49,300 and a transmembrane topology similar to that of other G protein-coupled receptors. Transient or stable expression of the cloned gene in host cells established specific ligand binding and functional activity characteristic of a D₁ dopamine receptor coupled to stimulation of adenylyl cyclase. Northern blot analysis and *in situ* hybridization

revealed that the messenger RNA for this receptor is most abundant in caudate, nucleus accumbens and olfactory tubercle, with little or no mRNA detectable in substantia nigra, liver, kidney, or heart. Several observations from this work in conjunction with results from other studies are consistent with the idea that other D₁ dopamine receptor subtypes may exist.

To isolate the gene for a D₁ dopamine receptor, a human retina complementary DNA library was screened at low stringency with an oligonucleotide probe derived from the second transmembrane segment (TMS) (amino acids 67–90) of the rat striatal D₂ dopamine receptor². A retinal library was chosen because dopamine is the predominant catecholamine in this tissue³ and there are abundant levels of both D₁ and D₂ receptors^{4,5}. A second TMS probe was used for screening because this region is highly conserved among catecholamine receptors. From this library, a clone (D233) was isolated that contained a 2.9-kilobase (kb) insert. D233 encoded an open reading frame that was highly homologous to other G protein-coupled receptors. But D233 seemed to be a partial clone lacking sequences homologous to the first TMS and amino-terminal region of these receptors.

The gene encoding this protein was judged to be intronless on the basis of the presence of a single hybridizing band on a Southern blot of human genomic DNA probed with a 0.7-kb *PvuII* restriction fragment of D233 at high stringency. At low stringency, however, several hybridizing bands were detected (data not shown). The *PvuII* fragment was then used to screen a human genomic DNA library. A 2.3-kb *EcoRI-XbaI* restriction fragment from one (HGL26) of three isolated clones possessed a nucleotide sequence extending beyond the 5' end of D233 and contained a likely site of translational initiation. Figure 1 shows the restriction map, nucleotide sequence (3,341 base pairs (bp)), and deduced amino-acid sequence derived from the overlapping sequences of genomic clone HGL26 and retinal cDNA clone D233. The open reading frame extends for 1,338 bp (446 amino acids).

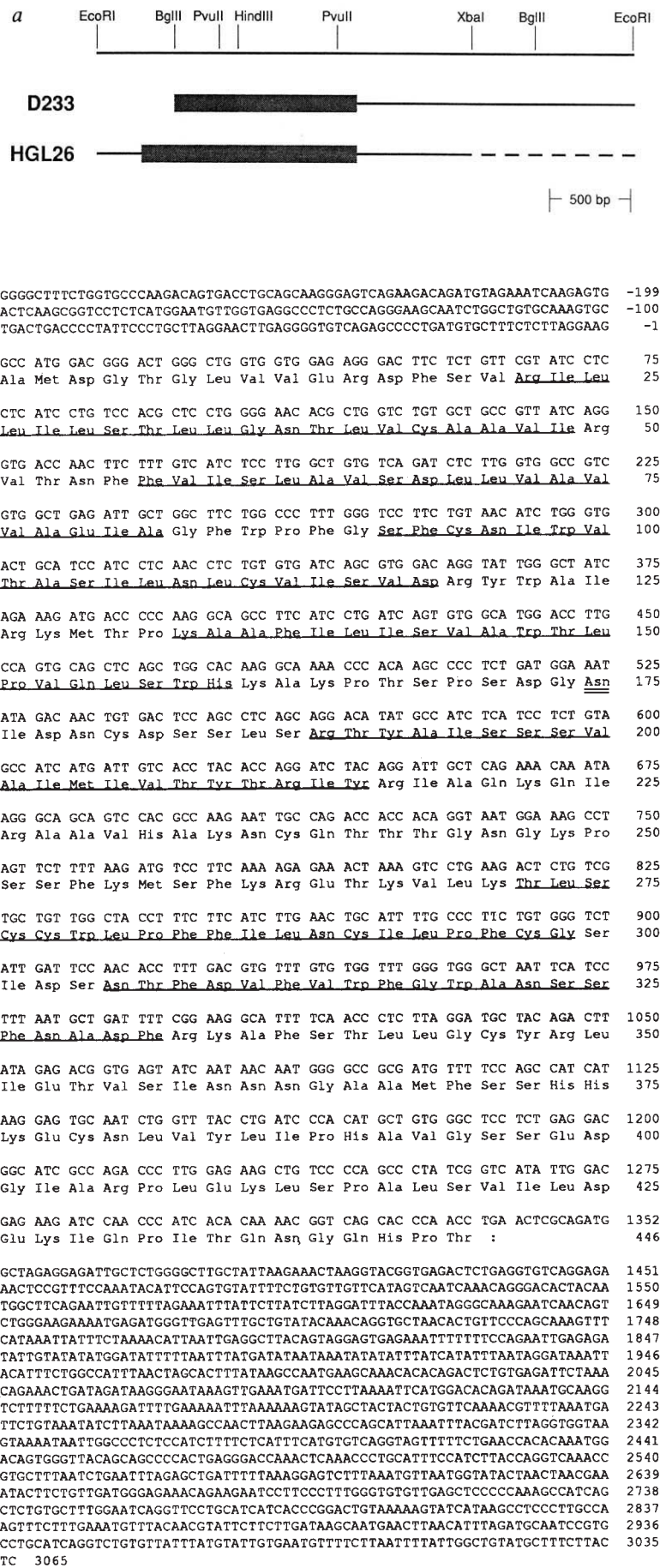
To identify the protein encoded by HGL26, the *EcoRI-XbaI* restriction fragment was subcloned into the expression vector pCMV5 (ref. 6) (pCMV5-D₁) and transiently transfected into COS-7 cells. Membranes prepared from these cells exhibited a 10- to 30-fold increase in specific binding of the D₁ antagonist [¹²⁵I]SCH23982, compared with cells transfected with an antisense construct. [¹²⁵I]SCH23982 bound with a dissociation constant (*K_d*) of 0.35 ± 0.02 nM, a value in good agreement with previous results obtained with rat striatal membranes⁷. Agonists and antagonists competed for the binding of [¹²⁵I]SCH23982 with an order of potency typical of binding to a D₁ dopamine receptor (Fig. 2a, b; Table 1)^{8,9}. Thus, the gene that we have isolated encodes a protein with the pharmacological characteristics expected for a D₁ dopamine receptor.

The D₁ receptor described here seems to couple selectively to adenylyl cyclase. In mouse Ltk⁻ cells transfected with pCMV5-D₁, dopamine stimulated membrane adenylyl cyclase activity roughly fivefold (Fig. 2c). This effect of dopamine was blocked by the D₁ antagonist SCH23982 but not by the D₂ antagonist raclopride. The D₁ partial agonist SKF38393 also stimulated enzyme activity, whereas the D₂ selective agonist quinpirole was ineffective. In contrast, dopamine failed to stimulate phosphatidylinositol (PtdIns) metabolism in COS-7 cells transfected with pCMV5-D₁, whereas adrenaline potently evoked this response in cells transfected with an α_1 -adrenergic receptor clone¹⁰ (data not shown).

The D₁ dopamine receptor encoded by HGL26 possesses several features characteristic of G protein-coupled receptors. Hydropathicity analysis of the protein sequence suggests the presence of seven TMS. Within these regions, the human D₁ dopamine receptor has a sequence identity of 40–43% with the human D₂ dopamine, α_2 -, β_1 - and β_2 -adrenergic, and 5-hydroxytryptamine 5HT_{1A} receptors. The most homologous regions shared by the human D₁ and D₂ receptors are found in TMS

FIG. 1 Restriction map (a) and nucleotide and deduced amino-acid sequence (b) of a human D₁ dopamine receptor clone. The sequence shown in b was obtained from D233, a cDNA clone from a human retina library, and HGL26, a human genomic clone. These two clones overlap as indicated in a. No differences were detected in the overlapping sequences of these clones. Restriction endonuclease sites are indicated. The 3' EcoRI site is an insertion site of the retinal cDNA clone into the λ arms. Stippled bars, coding region. Nucleotide sequence (b) is numbered in the 5' to 3' direction beginning with the first ATG of the open reading frame. Preceding bases are indicated by negative numbers. Nucleotides are numbered at the right-hand end of each line. The first ATG triplet downstream of nonsense codons found in frame with the rest of the coding block and having a Kozak consensus sequence³³ was assigned as the site of translational initiation. The deduced amino-acid sequence, shown below the nucleotide sequence, is numbered at the right-hand end of each line, beginning with the initiator methionine (Met). Hydrophobic segments representing putative membrane domains are underlined. Putative sites of N-linked glycosylation are indicated by double underlines. In the 3' untranslated region, several polyadenylation signals are found (beginning at bp 1,910, 2,288, 2,341, 2,907 and 3,040).

METHODS. Clone D233 was obtained by screening 2×10^6 recombinants from a human retina cDNA library in λ gt10 with a synthetic 72-base oligonucleotide probe derived from nucleotides 199–270 of a rat striatal D₂ dopamine receptor clone. The probe was purified on a denaturing polyacrylamide gel and labelled with ³²P at the 5' hydroxyl group using T4 polynucleotide kinase. Duplicate nitrocellulose filters were hybridized in $2 \times$ SSC buffer, $10 \times$ Denhardt's solution, 0.1% sodium pyrophosphate, 0.1% SDS, $50 \mu\text{g ml}^{-1}$ sheared salmon sperm DNA, and ³²P-labelled oligonucleotide ($2\text{--}10 \times 10^6$ c.p.m. ml^{-1}) at 42°C for 18 h. Filters were washed at 60°C in $1 \times$ SSC. A 0.7-kb PvuII restriction fragment from clone D233 was random-labelled with [³²P]dCTP and used to probe duplicate filters lifted from a human genomic library in EMBL3 (Clontech) under the following conditions: 50% formamide, $5 \times$ SSC, $5 \times$ Denhardt's, 25 mM sodium phosphate, 0.1% sodium pyrophosphate, $\sim 0.1\%$ SDS, $50 \mu\text{g ml}^{-1}$ sheared salmon sperm DNA, $100 \mu\text{g ml}^{-1}$ transfer RNA, 10% dextran sulphate, 42°C, 18 h. Filters were washed at 55°C in $1 \times$ SSC. Nucleotide sequencing was by Sanger dideoxy chain termination method using Sequenase (USB).



II and III (57% identity). Previous studies suggest that an Asp residue in TMS III and two Ser residues in TMS V, found in all known catecholamine receptors, interact with the amino and catechol hydroxyl groups, respectively, of catecholamines^{11,12}. These residues are conserved in the D₁ dopamine receptor (Asp 103, Ser 199, Ser 202). In addition, the sequence of the human D₁ receptor at the amino terminal of the third cytoplasmic loop (RIYRIAQKQIRRI in single-letter amino-acid code; amino acids 216–228) is very similar to the same region of the G_s-coupled human β_1 -adrenergic receptor (RVFREAQKQVKKI; amino acids 246–258). This intracellular loop may have an important role in coupling receptors to G proteins¹³. Other features of interest within this receptor are several potential sites of post-translational modification: two consensus sites¹⁴ for N-linked glycosylation (Asn 5 and Asn 175); a cysteine residue (Cys 347) in the carboxy terminus near TMS VII that is conserved in most G protein-coupled receptors and may be palmitoylated as in the β_2 -adrenergic receptor¹⁵ and rhodopsin¹⁶; and several consensus sites for regulatory phosphorylation¹⁷ in the predicted cytoplasmic domains.

The tissue distribution of D₁ dopamine receptor mRNA was examined by conventional northern blot analysis, polymerase chain reaction (PCR), and *in situ* hybridization. In northern blots of human tissues, mRNA for the receptor (4.2 kb) was most abundant in the caudate; somewhat lower levels were found in the cortex and cerebellum; no specific hybridization was detected in heart, liver, or kidney (data not shown). To detect receptor message that might be present at very low abundance, PCR was used to amplify a specific 490-bp fragment in the 3' untranslated region of the D₁ receptor message¹⁸. Although this technique is quite sensitive to low levels of message, high levels will be underestimated (see Fig. 3). A large amount of amplified product was obtained from human caudate RNA (Fig. 3a). The next greatest amount of product was obtained from frontal cortex, followed by hippocampus, cerebellum and ventral tegmental area. No detectable product was amplified from substantia nigra, kidney, heart, or liver, even though, with the exception of liver, all these tissues possess D₁ binding sites^{19–21}.

We analysed in more detail the distribution of D₁ receptor message in the central nervous system by *in situ* hybridization using a partial cDNA clone encoding the rat homologue of the human receptor. Areas with high message levels were striatum, nucleus accumbens and olfactory tubercle (Fig. 3c). No signal was detectable in substantia nigra (Fig. 3d). By contrast, hybridization of serial sections with a probe specific for D₂

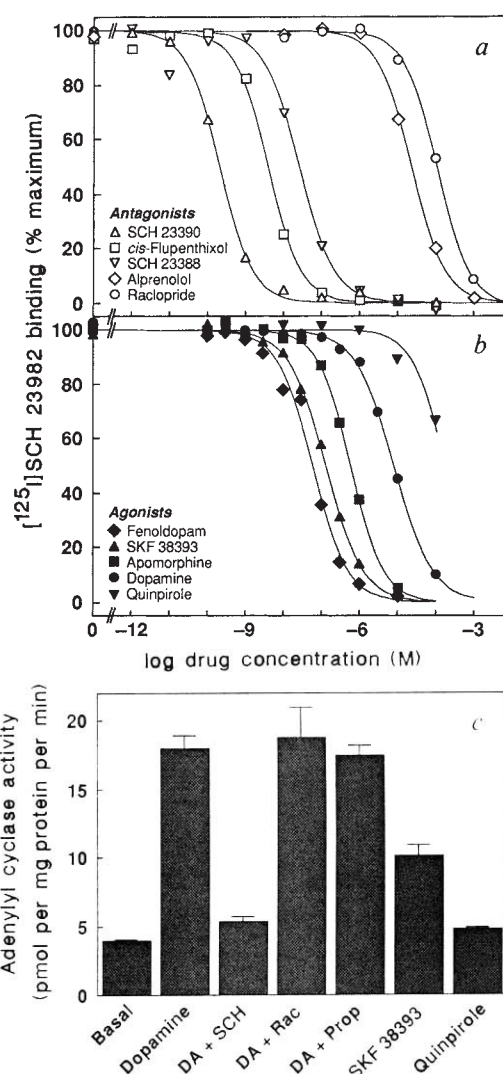


FIG. 2 *a* and *b*, Binding properties of the selective D₁ dopamine receptor antagonist [¹²⁵I]SCH23982 to membranes of COS-7 cells expressing clone HGL26. Various catecholamine antagonists (*a*) or agonists (*b*) were tested for their ability to compete for the binding of [¹²⁵I]SCH23982 to membranes of African green monkey kidney (COS-7) cells transfected with a pCMV5 construct of clone HGL26. *c*, Stimulation by dopamine of adenylyl cyclase activity in membranes prepared from mouse fibroblasts expressing the D₁ dopamine receptor. Drug concentrations: dopamine (DA), 100 μ M; SCH23390 (SCH), raclopride (Rac), propranolol (Prop), SKF38393 and quinpirole, 1 μ M each.

METHODS. *a* and *b*, A 2.3-kb *EcoRI*–*XbaI* restriction fragment of clone HGL26 was subcloned into the same sites of the expression vector pCMV5 (ref. 6). This pCMV5–D₁ construct was used to transfect³⁴ COS-7 cells for transient expression of the D₁ receptor. Binding assays were performed on crude membrane suspensions prepared from the cells. Saturation assays indicated that the apparent equilibrium dissociation constant (K_d) for [¹²⁵I]SCH23982 was 0.35 ± 0.02 nM ($n=6$); non-specific binding was determined with 1 μ M *cis*-flupenthixol. Competition assays were versus 0.3 nM [¹²⁵I]SCH23982 with varying concentrations of agonists or antagonists. Incubations were terminated by rapid vacuum filtration. Nonspecific binding typically represented 3–10% of total binding. Data are average values of duplicate determinations from representative experiments and were modelled by nonlinear least square curve fitting by computer using a one-site model³². K_d values calculated from these and other data are given in Table 1. *c*, Mouse fibroblast [L–M(TK⁺)] cells stably expressing the D₁ dopamine receptor were obtained by cotransfection of pCMV5–D₁ with the selection plasmid pRSVNeo (ref. 34) followed by selection with G418. Crude membrane suspensions from the cells were prepared and resuspended to ~ 0.2 mg protein ml⁻¹ in 75 mM Tris–HCl, pH 7.2, 5 mM MgCl₂, 3 mM EDTA. Adenylyl cyclase activity was measured by the method of Salomon *et al.*³⁵ as modified by Bouvier *et al.*³⁶, except that 2 mM MgCl₂ was used in the assay.

TABLE 1 K_d values of ligands for [¹²⁵I]SCH23982 binding sites

Antagonists	K_d (nM)	Agonists	K_d (nM)
SCH23390	0.11	fenoldopam	20
<i>cis</i> -piflutixol	0.20	SKF38393	87
(+) butaclamol	0.90	(-) apomorphine	210
<i>cis</i> -flupenthixol	1.6	dopamine	2,500
SCH23388	22	quinpirole	14,000
ketanserin	190	serotonin	>15,000
piperone	220	adrenaline	>55,000
yohimbine	1,800		
prazosin	3,000		
(-) butaclamol	4,900		
alprenolol	16,000		
raclopride	>72,000		

Varying amounts of antagonists and agonists were used to compete for [¹²⁵I]SCH23982 binding to membranes prepared from COS-7 cells transfected with the *EcoRI*–*XbaI* restriction fragment of clone HGL26 in the expression vector pCMV5. Experiments were conducted as in Fig. 2. Apparent equilibrium dissociation constants (K_d) were determined by computer with a nonlinear regression program describing the interaction of ligands with a single class of binding sites³². Values represent the means of two to five independent experimental determinations whose pK_d values were within 10%.

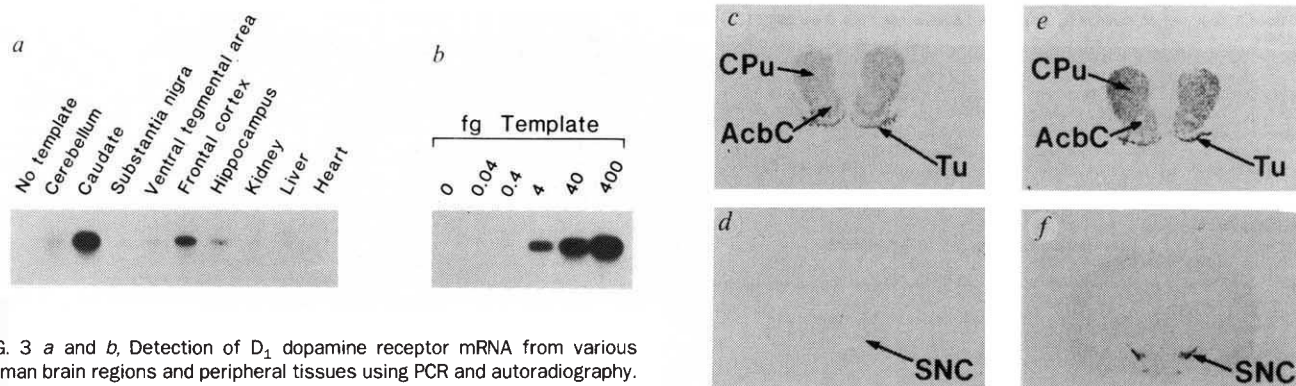


FIG. 3 *a* and *b*, Detection of D₁ dopamine receptor mRNA from various human brain regions and peripheral tissues using PCR and autoradiography. The most intense signal was observed in caudate. Less intense signals were observed in frontal cortex, hippocampus, cerebellum and ventral tegmental area. No signal was observed in substantia nigra, kidney, liver, or heart. *b*, Standard curve (fg, femtogram). *c*–*f*, Localization of D₁ receptor mRNA in rat brain by *in situ* hybridization as shown by X-ray film images of rat brain coronal sections. *c* and *d*, Hybridization with ³⁵S-labelled antisense D₁ receptor probe. Note the intense hybridization signal overlying the striatum (CPu), nucleus accumbens (AcbC), and olfactory tubercle (Tu). *d* shows that no specific labelling was observed over the substantia nigra (SNC) or ventral tegmental area. *e* and *f*, Hybridization with ³⁵S-labelled antisense D₂ receptor probe. Note the intense hybridization signal over the CPu, AcbC, Tu, and SNC. Hybridization of serial sections with ³⁵S-labelled sense-strand control probes from the D₁ or D₂ receptor clones resulted in uniform background labelling.

METHODS. *a*, Total RNA was isolated from human post-mortem tissue using the guanidinium isothiocyanate-caesium chloride method³⁷. Using this total RNA (2 µg) an 18-mer (ATGAACATTTAGAACTCT) primer directed to the 3' untranslated region of the human D₁ dopamine receptor was used to synthesize first-strand cDNA using AMV-reverse transcriptase (4 U) (Bethesda Research Laboratory) under conditions previously described³⁸. Polymerase from *Thermus aquaticus* (1 U), the appropriate buffer, 20 µCi [³²P]dCTP, and two new 18-mer primers (CCAAATACATTCCAGTGT as 5' primer; AAATGTGGTGTAAAGTC as 3' primer) were then added to the reaction. These components were subjected to 25 cycles of amplification (94 °C, 2 min; 45 °C, 2 min; 72 °C, 2 min) to amplify a 490-bp segment of this first-strand cDNA template. An aliquot of each reaction was electrophoresed on a 1.5% agarose gel, transferred to a nylon membrane, washed, and exposed to X-ray film. To test for the presence of contaminating genomic DNA in the RNA from these tissues, a reaction was performed as described above but without the addition of AMV-reverse transcriptase. In

this experiment, no detectable bands were amplified from RNA (2 µg) from each of these tissues (data not shown). Furthermore, primers alone without any RNA template did not give rise to any detectable product (lane 1 of *a*). *b*, A standard curve was developed to test the linearity of the PCR amplification. The 2.3-kb *EcoRI*–*XbaI* restriction fragment of clone HGL26 was subcloned into pBluescript II SK+ (Stratagene). This construct was digested with *XbaI*, and RNA was synthesized using T7 RNA polymerase. The resulting product was digested with DNase I, diluted, and 400, 40, 4, 0.4, and 0.04 fg of this template RNA were subjected to the procedure described above. Aliquots from each reaction were applied to a Whatman DE-81 membrane, washed extensively in 0.5 M phosphate buffer, pH 6.5, and counted by liquid scintillation spectrometry. The standard curve indicated that the amount of amplified product was proportional to the amount of template added. But this relationship was no longer linear at higher levels of template where the amount of enzyme and primers in the reaction presumably become limiting. Thus, this method is quite sensitive for very low abundance messages but is limited by the amount of product formed when template is very abundant. Hence, the differences observed between tissues are likely to be larger than those indicated by the autoradiogram in *a*. *c*–*f*, A rat striatal cDNA library was screened (1.5 × 10⁶ recombinants) using the 0.7-kb *PvuII* restriction fragment of clone D233 following methods similar to those described for screening of the human genomic library (see Fig. 1). A 0.45-kb *EcoRI*–*Clal* restriction fragment from an isolated clone (RSL25) and a 1.5-kb *SacI*–*BamHI* restriction fragment from the rat striatal D₂ receptor cDNA (ref. 2) were subcloned into pBluescript (Stratagene). ³⁵S-labelled antisense- or sense-strand RNA probes were prepared by *in vitro* transcription and hybridized with 4% paraformaldehyde-fixed coronal rat brain sections (15 µm) as previously described³⁹.

dopamine receptor mRNA revealed the expected²² high level of signal in substantia nigra zona compacta (Fig. 3*f*) as well as striatum and nucleus accumbens (Fig. 3*e*). The lack of D₁ receptor mRNA in substantia nigra, as determined by both PCR and *in situ* hybridization, is consistent with previous lesion experiments suggesting that the D₁ receptors in this area are on terminals of the striatonigral pathway neurons²³.

Our pharmacological and functional characterization of genomic clone HGL26 demonstrates that this clone encodes a D₁ dopamine receptor coupled to stimulation of adenylyl cyclase that is similar to the D₁ receptor previously characterized in a variety of tissues^{24,25}. Other classes of G protein-coupled receptors such as β-, α₁- and α₂-adrenergic, muscarinic and serotonergic receptors, have recently been found to include multiple subtypes distinguished by their tissue distribution, pharmacological properties, or transduction mechanisms²⁶. Several observations from this work as well as results from other studies²⁷ are consistent with the idea that subtypes of D₁ dopamine receptors may also exist. For example, dopamine reportedly stimulates PtdIns turnover through a D₁ receptor in the striatum²⁸ and in *Xenopus* oocytes injected with mRNA from rat striatum²⁹. As dopamine does not stimulate PtdIns metabolism in cells expressing the D₁ receptor described here, striatum may also contain a different D₁ receptor linked to PtdIns turnover. In addition, previous studies have documented the presence of D₁ receptors in the kidney coupled to either adenylyl cyclase³⁰ or phospholipase C (ref. 31) that exhibit

pharmacological differences from CNS D₁ receptors²⁰. Our results from northern blot and PCR analysis of message demonstrate a lack of mRNA for the cloned D₁ receptor in the kidney. These findings, coupled with our observation of multiple hybridizing bands on Southern blot analysis at low stringency, are consistent with the existence of additional D₁ dopamine receptor subtypes. Cloning of the gene for a D₁ dopamine receptor will facilitate investigation of this possible heterogeneity as well as the molecular processes involved in activation and regulation of this receptor, the functional interaction of D₁ and D₂ receptors, and the potential role of these receptors in pathophysiological processes. □

Received 25 April; accepted 14 June 1990.

1. Keibian, J. & Calne, D. *Nature* **277**, 93–96 (1979).
2. Bunzow, J. et al. *Nature* **336**, 783–787 (1988).
3. Ehinger, B. *Prog. Retinal Res.* **2**, 213–232 (1983).
4. Gredal, O., Parkinson, D. & Nielsen, M. *Eur. J. Pharmacol.* **137**, 241–245 (1987).
5. McGonigle, P., Wax, M. & Molinoff, P. *Invest. Ophthalmol. Vis. Sci.* **29**, 687–694 (1988).
6. Andersson, S., Davis, D., Dahlback, H., Jönvall, H. & Russell, D. *J. Biol. Chem.* **264**, 8222–8229 (1989).
7. Sidhu, A., van Oene, J., Dandridge, P., Kaiser, C. & Keibian, J. *Eur. J. Pharmacol.* **128**, 213–220 (1986).
8. Billard, W., Ruperto, V., Crosby, G., Iorio, L. & Barnett, A. *Life Sci.* **35**, 1885–1893 (1984).
9. Andersen, P., Gronwald, F. & Jansen, A. *Life Sci.* **37**, 1971–1983 (1985).
10. Cotecchia, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7159–7163 (1988).
11. Dixon, R. et al. *Nature* **326**, 73–77 (1987).
12. Strader, C., Candelore, M., Hill, W., Sigal, I. & Dixon, R. *J. Biol. Chem.* **264**, 13572–13578 (1989).
13. O'Dowd, B. et al. *J. Biol. Chem.* **263**, 15985–15992 (1988).
14. Hubbard, S. & Ivatt, R. A. *Rev. Biochem.* **50**, 555–583 (1981).

15. O'Dowd, B., Hnatowich, M., Caron, M. G., Lefkowitz, R. J. & Bouvier, M. *J. Biol. Chem.* **264**, 7564–7569 (1989).
16. Ovchinnikov, Y., Abdulaev, N. & Bogachuk, A. *FEBS Lett.* **230**, 1–5 (1988).
17. Caron, M. et al. in *G Proteins* 295–316 (Academic, Orlando, Florida, 1990).
18. Wang, A., Doyle, M. & Mark, D. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9717–9721 (1989).
19. Dawson, T., Barone, P., Sidhu, A., Wamsley, J. & Chase, T. *Neuroscience* **26**, 83–100 (1988).
20. Felder, C., Jose, P. & Axelrod, J. *J. Pharmac. exp. Ther.* **248**, 171–175 (1989).
21. Sandrini, M., Benelli, A. & Baraldi, M. *Life Sci.* **35**, 1839–1849 (1984).
22. Meador-Woodruff, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7625–7628 (1989).
23. Spano, P., Trabucchi, M. & Di Chiana, G. *Science* **196**, 1343–1345 (1977).
24. Creese, I., Morrow, L., Leff, S., Sibley, D. & Hamblin, M. *Int. Rev. Neurobiol.* **23**, 255–301 (1982).
25. Niznik, H. *Molec. cell Endocrinol.* **54**, 1–22 (1987).
26. Schofield, P., Shivers, B. & Seeburg, P. *Trends Neurosci.* **13**, 8–11 (1990).
27. Andersen, P. et al. *Trends Pharmacol. Sci.* **11**, 231–236 (1990).
28. Undie, A. & Friedman, E. *Ninth Int. Washington Spring Symp. Biol. Cell. Transducing Signals* 261 (1989).
29. Mahan, L., Burch, R., Monsma, F. & Sibley, D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2196–2200 (1990).
30. Felder, R., Blecher, M., Calcagno, P. & Jose, P. *Am. J. Physiol.* **247**, F499–F505 (1984).
31. Felder, C., Blecher, M. & Jose, P. *J. Biol. Chem.* **264**, 8739–8745 (1989).
32. DeLean, A., Hancock, A. & Lefkowitz, R. *Molec. Pharmacol.* **21**, 5–16 (1982).
33. Zokak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).
34. Cullen, B. *Meth. Enzym.* **152**, 684–704 (1987).
35. Salomon, Y., Londos, C. & Rodbell, M. *Analyt. Biochem.* **58**, 541–548 (1974).
36. Bouvier, M. et al. *Molec. Pharmacol.* **33**, 133–139 (1988).
37. Chirgwin, J., Przybyla, A., MacDonald, R. & Rutter, W. *Biochemistry* **18**, 5294–5299 (1979).
38. Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
39. Freneau, R. Jr., Lundblad, J., Pritchett, D., Wilcox, J. & Roberts, J. *Science* **234**, 1265–1269 (1986).

ACKNOWLEDGEMENTS. We thank J. Nathans for the human retina cDNA library, M. Ehrlich for the rat striatal cDNA library, M. Hnatowich for assistance with sequence comparisons, J. Lomasney for providing human brain tissue, O. Civelli for a cDNA clone encoding a rat striatal D₂ dopamine receptor, D. Addison for assistance with manuscript preparation, and V. Clack and N. Godinot for technical excellence. This work was supported in part by the NIH.

Cloning and expression of human and rat D₁ dopamine receptors

Qun-Yong Zhou^{*†}, David K. Grandy^{*}, Lisa Thambi^{*‡},
Jake A. Kushner^{*}, Hubert H. M. Van Tol^{*§},
Roger Cone^{*}, David Pribnow^{||}, John Salon^{*},
James R. Bunzow^{*} & Olivier Civelli^{*||}

^{*} Vollum Institute for Advanced Biomedical Research, [†] Department of Biochemistry and Molecular Biology, [‡] Department of Pharmacology and ^{||} Department of Cell Biology and Anatomy, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201, USA

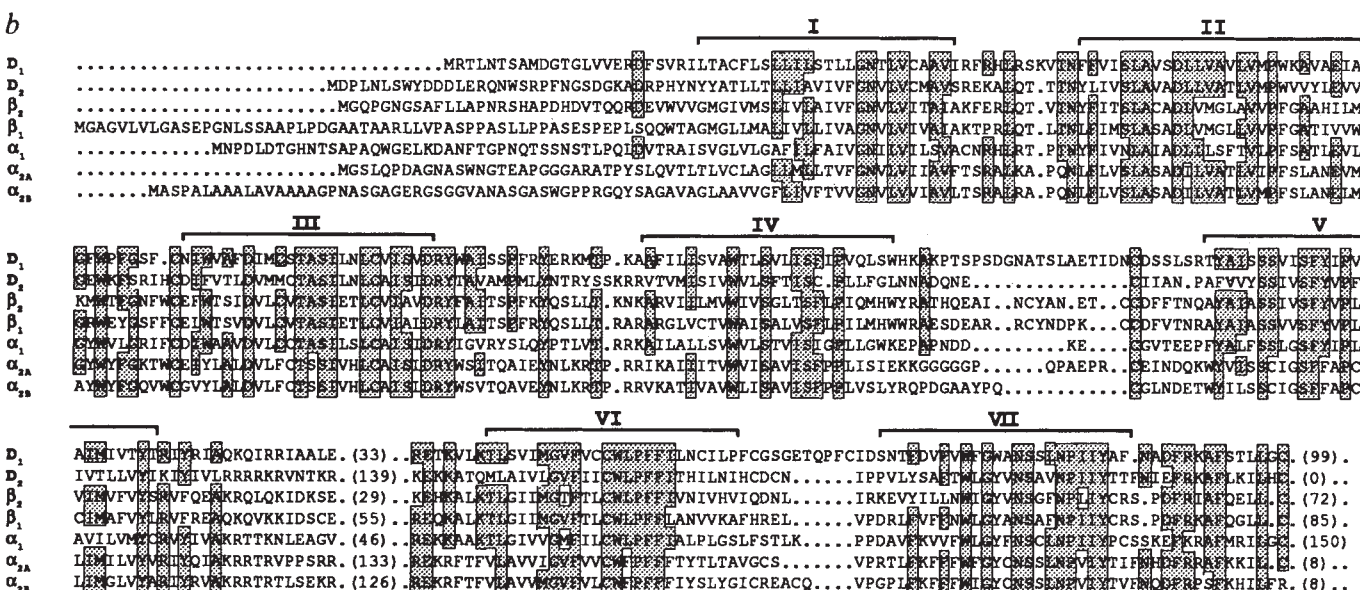
[§] Present address: Department of Pharmacology, Medical Sciences Building, University of Toronto, Toronto, Ontario M5S 1A8, Canada

[¶] To whom correspondence should be addressed

THE importance of the dopaminergic system in brain function has been emphasized by its association with neurological and psychiatric disorders such as Parkinson's disease and schizophrenia. On the basis of their biochemical and pharmacological characteristics, dopamine receptors are classified into D₁ and D₂ subtypes^{1,2}. As the most abundant dopamine receptor in the central nervous system, D₁ receptors seem to mediate some behavioural responses³, modulate activity of D₂ dopamine receptors^{4,5}, and regulate neuron growth and differentiation⁶. The D₂ dopamine receptor has been cloned by low-stringency screening⁷. We report here the cloning

of human and rat D₁ dopamine receptors by applying an approach based on the polymerase chain reaction⁸. The cloned human D₁ dopamine receptor has been characterized on the basis of four criteria: the deduced amino-acid sequence, which reveals that it is a G protein-coupled receptor; the tissue distribution of its messenger RNA, which is compatible with that of the D₁ dopamine receptor; its pharmacological profile when transfected into COS-7 cells; and its ability to stimulate the accumulation of cyclic AMP in human 293 cells.

The D₁ and D₂ dopamine receptors are G protein-coupled receptors that stimulate and inhibit adenylyl cyclase, respectively^{1,2}. Their structures should therefore contain the putative seven transmembrane domains common to G protein-coupled receptors⁷. The rat D₂ receptor was cloned by low-stringency screening using the hamster β_2 -adrenergic receptor as a probe⁷. To clone the D₁ receptor, we applied a different approach. A set of degenerate oligonucleotide primers were designed based on the nucleotide sequences of known catecholamine receptors and some other G protein-coupled receptors^{9,10}. Two primers, corresponding to putative receptor transmembrane domains III and VI, were used in a polymerase chain reaction (PCR)⁸. Rat striatum complementary DNA was chosen as the template because high levels of D₁ dopamine receptor have been found in this tissue¹¹. Deletion studies have shown that the third cytoplasmic loop is crucial for G-protein coupling¹². As the three cloned β -adrenergic receptors that couple to G_s proteins have putative third cytoplasmic loops of 52–78 amino acids¹³, we hypothesized that the third cytoplasmic loop of the dopamine D₁ receptor might be in a similar size range. Therefore, the PCR



products were size-fractionated and products ranging from 450–700 base pairs (bp) were subcloned into M13 and subjected to direct sequencing. Of 24 PCR products analysed, D₂ dopamine⁷, α_{2B} -adrenergic¹⁴ and five sequences representing potentially new G protein-coupled receptors were obtained. One of these clones, R213, had several interesting structural features. It had a higher degree of amino-acid similarity with known catecholamine receptors as compared with other G protein-coupled receptors; in the putative fifth transmembrane domain it contained two serine residues that were thought to be specific to receptors binding catecholamines¹⁵; and it had a putative third cytoplasmic loop similar in size and sequence to that of the β -adrenergic receptors¹³.

The PCR-generated clone R213 was used as a probe to screen a rat striatum cDNA library. One positive clone was identified and sequenced. Although not full-length, it allowed us to describe most of the rat coding sequence (Fig. 1). As most catecholamine receptors lack introns in their coding regions¹⁶ and as our preliminary human genomic analysis indicated the absence of introns in this gene (D. K. Grandy *et al.*, unpublished observations), we screened a human genomic library. Eight positive signals were obtained. One clone, HGR213-1, was further characterized and a 3.0-kilobase (kb) *EcoRI*–*SacI* fragment spanning the whole coding region was subcloned and sequenced.

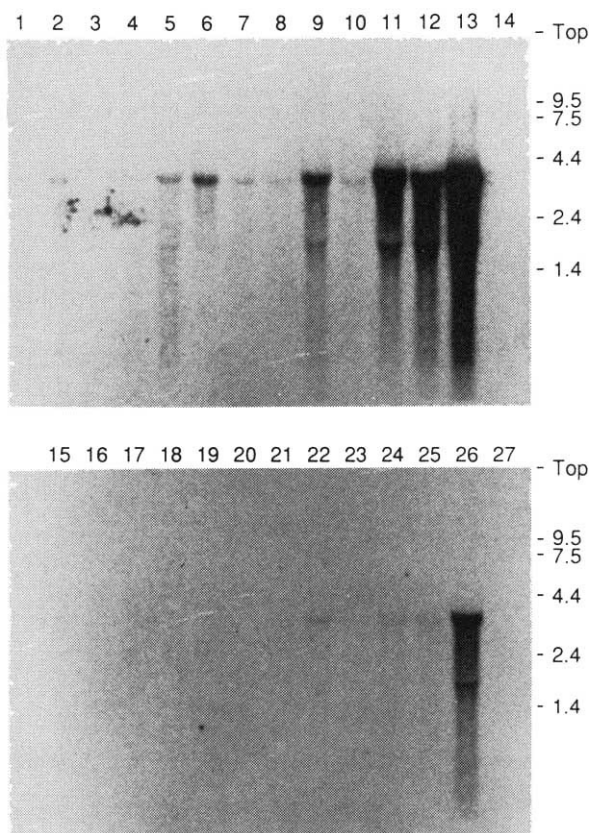


FIG. 2 Northern blot analysis of R213 transcripts in rat brain regions, pituitary and peripheral tissues. Northern blot analysis was performed as previously described, except random-primed R213 was used as hybridization probe⁷. Each lane contained 20 μ g total RNA. Numbers on the right, RNA size markers (kb) (BRL). Lane 1, olfactory bulb; 2, hippocampus; 3, cerebellum; 4, posterior cortex; 5, anterior cortex; 6, thalamus; 7, hypothalamus; 8, medulla; 9, amygdala; 10, mesencephalon; 11, septum; 12, posterior basal ganglia; 13, anterior basal ganglia; 14, neurointermediate lobe of pituitary; 15, muscle; 16, ventricle; 17, atrium; 18, lung; 19, adrenal; 20, kidney; 21, liver; 22, pineal; 23, anterior lobe of pituitary; 24, hypothalamus; 25, mesencephalon; 26, posterior basal ganglia; 27, neurointermediate lobe of pituitary.

Figure 1 shows the nucleotide sequence of clone HGR213-1. The longest open reading frame codes for a 446-amino-acid protein (relative molecular mass 49,296 ($M_r \sim 49K$)). This M_r is similar to the reported value of the deglycosylated form of dopamine D₁ receptor as determined by SDS-PAGE¹⁷. Like most adrenergic receptors, but unlike the dopamine D₂ receptor¹⁸, HGR213-1 has no intron in its coding sequence. There are two potential in-frame initiation sites. Considering the unique potential *N*-linked glycosylation site in the *N*-terminus, the initiation site shown in Fig. 1 is probably the one that is used. There is another potential *N*-linked glycosylation site in the second extracellular loop.

Hydrophobicity analysis of HGR213-1 revealed seven stretches of hydrophobic amino acids that could represent transmembrane domains (data not shown). Comparison of the deduced amino-acid sequence of HGR213-1 with that of other catecholamine receptors shows that the greatest similarity exists in the putative transmembrane domains, where the amino-acid

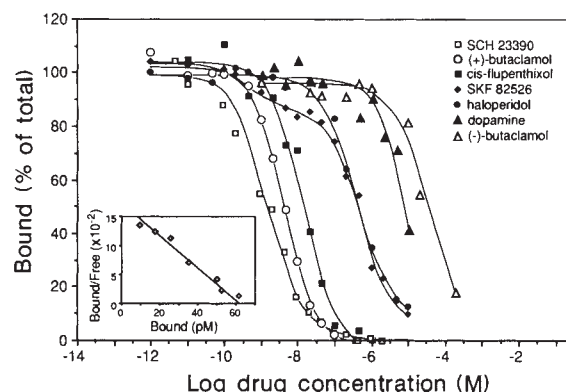
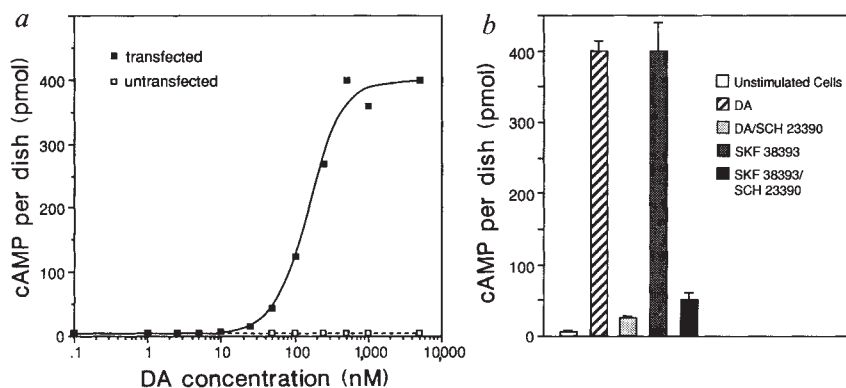


FIG. 3 Binding of [³H]SCH23390 to membranes prepared from COS-7 cells transfected with HGR213-1. Representative curves show the competitive inhibition of [³H]SCH23390 specific binding by different drugs. The inset shows a Scatchard transformation of saturation binding. The average inhibition constant, K_i , values from three independent experiments were: SCH23390, 0.4 nM; (+) butaclamol, 2.1 nM; *cis*-flupenthixol, 5.6 nM; haloperidol, 203 nM; dopamine, 2.3 μ M; (–)butaclamol, 19 μ M and SKF82526, 0.2 nM (high affinity) and 150 nM (low affinity). In the Scatchard plot shown, the dissociation constant K_d and the maximal number of binding sites, B_{max} , values for membranes prepared from transfected COS-7 cells were 0.3 nM and 2 pmol per mg protein, respectively.

METHODS. The 3.0-kb *EcoRI*–*SacI* fragment of HGR213-1 was inserted between the unique *HindIII* and *BamHI* sites of eukaryotic expression vector pBC12B1 (ref. 27). A modified calcium phosphate method²⁸ was used for the transfection of COS-7 cells. About 45 μ g plasmid DNA were used for each large 150 mm plate. At 48 h after transfection, cells were rinsed with TEM buffer (25 mM Tris buffer, pH 7.4, 6 mM MgCl₂, 1 mM EDTA) and scraped off plates. Membranes were prepared by homogenizing cells with a ConTorque homogenizer at 4 °C in TEM buffer. The homogenate was centrifuged at 800g for 10 min and the pellet again homogenized and centrifuged. Supernatants were pooled and centrifuged at 100,000g for 1 h. The pellet was then resuspended in TEM buffer at appropriate protein concentration and stored in small aliquots at –70 °C. Binding assays were performed in duplicate in 500 μ l, containing 50 mM Tris buffer, pH 7.4, 0.9% NaCl, 0.025% ascorbic acid, 0.001% bovine serum albumin [³H]SCH23390 (Amersham, 69 Ci mmol^{–1}) and tested drugs. In all competition binding assays, 0.7 nM [³H]SCH23390 was inhibited by various concentrations of unlabelled drugs. Binding was initiated by the addition of membrane preparation (20–30 μ g protein) and carried out at 30 °C for 1 h. Nonspecific binding was defined in the presence of 10 μ M (+) Butaclamol. Samples were filtered through glass fibre filters (Schleicher and Schuell No. 32) and washed three times with 4 ml ice-cold 10 mM Tris buffer, pH 7.4. The radioactivity retained on the filter was counted using a Beckman LS6800 scintillation counter. The 50% inhibitory concentration values (IC_{50}) calculated from the curves were converted to K_i values as described⁷. Inhibition was fitted best by assuming the existence of only one class of binding site, except in the case of inhibition by the agonist SKF82526, which was best fitted by assuming the presence of two classes of binding sites. A LIGAND computer program was used for data analysis and curve fitting.

FIG. 4 *a*, Dopamine-induced cAMP accumulation in human embryonic kidney 293 cells transfected with HGR213-1. Intracellular cAMP was measured as a function of dopamine (DA) concentration after transient expression of HGR213-1. Triplicate plates were analysed for each point. The half-maximal stimulation concentration of dopamine (EC₅₀) of the curve shown is 154 nM. *b*, Stimulation of cAMP accumulation in 293 cells by dopamine and SKF38393 and the antagonizing effect of SCH23390. Cyclic AMP production in 293 cells was stimulated by the agonists dopamine (125 nM) and SKF38393 (250 nM) and antagonized by SCH23390 (500 nM).



METHODS. Exponential growing human 293 cells (in 60-mm dishes) were transfected with HGR213-1 expression plasmid DNA (5 μ g) in PBC12BI (ref. 27) using a modified calcium phosphate method²⁹. The dishes were rinsed twice with DMEM plus 10% FCS after 18 h. Two days later, the plates were rinsed twice with DMEM containing 1 mg ml⁻¹ BSA and 0.5 mM IBMX (3-isobutyl-1-methylxanthine). The cells were then incubated for 45 min at 37 °C in the same medium containing various drugs. After aspiration of the medium, cells were washed twice with ice-cold Hank's buffered saline and lysed with 1 ml of 60%

ethanol. The cell debris was collected and pelleted and the supernatants lyophilized. The resulting pellets were resuspended in water and cAMP in each sample was quantitated using an assay method (Amersham) that measures the ability of cAMP in the sample to displace [8-³H] cAMP from a high affinity cAMP binding protein³⁰. The values obtained are normalized for the number of cells on a 60-mm dish ($\sim 5 \times 10^5$ cells in *a* and 10^6 cells in *b*).

identities are as follows: 44% with human D₂ (ref. 18); 42% with human β_2 (ref. 19); 43% with human β_1 (ref. 20); 41% with hamster α_1 (ref. 21); 42% with human α_{2A} (ref. 22) and 40% with human α_{2B} (ref. 14). The overall degree of identity between HGR213-1 and D₂ receptors is about the same as between HGR213-1 and adrenergic receptors. Asp 79 and Asp 113 in the β_2 -adrenergic receptor, which possibly act as counterions for the positively charged catecholamine²³, are present at corresponding positions in HGR213-1. Furthermore, the size and sequence of its third cytoplasmic loop and C terminus of HGR213-1 are similar to that of β -adrenergic receptors. This suggested to us that this new receptor might be coupled to G_s (ref. 12). But the absence of a potentially important glutamic acid residue²⁴, which is conserved in the third transmembrane domains of all three cloned β -adrenergic receptors¹³ indicated that HGR213-1 probably was not a β -adrenergic-like receptor. On the basis of these structural features, we hypothesized that HGR213-1 could encode a dopamine D₁ receptor. In addition, there are two consensus sequences (residues 133–136, 265–268) for cAMP-dependent protein kinase phosphorylation⁹, and the many serines and threonines in the cytoplasmic loops and the relatively long C terminus could be potential protein kinase C (ref. 25) or receptor kinase²⁶ phosphorylation sites.

As a step towards identifying HGR213-1, the tissue distribution of its transcript was examined by northern blot analysis. A messenger RNA of ~ 4 kb was found in many rat brain regions with the highest level of expression in the basal ganglia (Fig. 2). HGR213-1 mRNA was undetectable in the pituitary and in the peripheral tissues we tested. This pattern of HGR213-1 messenger distribution in the central nervous system and pituitary is consistent with that of the dopamine D₁ receptor as determined by autoradiography and binding studies¹¹.

To further investigate its identity, HGR213-1 was transiently expressed in eukaryotic cells. The 3.0-kb *EcoRI*–*SacI* fragment of HGR213-1 was inserted into eukaryotic expression vector PBC12I (ref. 27) and transfected into monkey kidney COS-7 cells. As its structural features and mRNA tissue distribution suggested that HGR213-1 might encode a dopamine D₁ receptor, membranes from transfected COS-7 cells were tested for their ability to bind to the D₁ selective antagonist [³H]SCH23390. Untransfected COS-7 cells showed no specific binding of [³H]SCH23390 (data not shown). Binding of [³H]SCH23390 to membranes prepared from transfected COS-7 cells was saturable with a dissociation constant (*K_d*) of 0.3 nM (Fig. 3 inset). This value agrees well with both the reported value²⁸ and the value observed in parallel experiments with rat striatal membranes (data not shown). Figure 3 shows competition curves of various ligands with [³H] SCH23390. The D₁ selective antagonist

SCH23390 and agonist SKF82526 were most potent, and the D₂ selective antagonist haloperidol was virtually inactive. The rank order of ligand potency was: SCH23390 > (+)butaclamol > flupenthixol > haloperidol. This pharmacological profile explicitly identifies the binding site as that of a dopamine D₁ receptor.

To demonstrate the HGR213-1 encodes a functional dopamine D₁ receptor we examined its ability to couple dopamine binding to activation of adenylyl cyclase. Human embryonic kidney 293 cells transiently expressing HGR213-1 were tested for their ability to respond to dopamine. When exposed to dopamine, untransfected cells showed no elevation of cAMP (Fig. 4*a*). In contrast, transfected cells displayed a concentration-dependent and saturable increase of intracellular cAMP levels with a half-maximal stimulation concentration (EC₅₀) of about 125 nM (Fig. 4*a*). This value is comparable to the previously reported value²⁸. SKF38393, a selective D₁ agonist, had a similar effect on the intracellular cAMP production and the stimulatory effects of both dopamine and SKF38393 were blocked by SCH23390 (Fig. 4*b*). These results indicated that the cloned D₁ receptor could couple positively to adenylyl cyclase.

On the basis of the above results we conclude that HGR213-1 encodes a human D₁ dopamine receptor. The successful cloning of the human D₁ dopamine receptor provides a new tool to study the regulation and function of this receptor. Moreover, the availability of both D₁ and D₂ dopamine receptor clones, which both bind to dopamine but couple to distinct effectors, should provide us with a new approach to address the complex interactions between these receptors. □

Received 26 April; accepted 15 June 1990.

1. Kebabian, J. W. & Calne, D. B. *Nature* **277**, 93–96 (1979).
2. Creese, I., Sibley, D. R., Hamblin, M. W. & Leff, S. E. A. *Rev. Neurosci.* **6**, 43–71 (1983).
3. Clark, D. & White, F. J. *Synapse* **1**, P347–P388 (1987).
4. Walters, J. R., Bergstrom, D. A., Carlson, J. H., Chase, T. N. & Braun, A. R. *Science* **236**, 719–722 (1987).
5. Waddington, J. L. *J. Psychopharmacol.* **3**, 54–63 (1989).
6. Lankford, K. L., DeMello, F. G. & Klein, W. L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4567–4571 (1988).
7. Bunzow, J. R. *et al. Nature* **336**, 783–787 (1988).
8. Saiki, R. K. *et al. Science* **239**, 487–491 (1988).
9. Dohman, H. G., Caron, M. G. & Lefkowitz, R. J. *Biochemistry* **26**, 2657–2664 (1987).
10. Libert, F. *et al. Science* **244**, 569–572 (1989).
11. Boyson, S. J., McGonigle, P. & Molinoff, P. B. *J. Neurosci.* **6**, 3177–3188 (1986).
12. O'Dowd, B. F. *et al. J. Biol. Chem.* **263**, 15985–15992 (1988).
13. Bunzow, J. R. *et al. Science* **245**, 1118–1121 (1989).
14. Regan, J. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6301–6305 (1988).
15. Strader, C. D., Candelore, M. R., Hill, W. S., Sigal, I. S. & Dixon, R. A. F. *J. Biol. Chem.* **264**, 13572–13578 (1989).
16. Lefkowitz, R. J. *et al. Cold Spring Harb. Symp. quant. Biol.* **53**, 507–514 (1988).
17. Jarvie, K. R., Booth, G., Brown, E. M. & Niznik, H. B. *Molec. Pharmacol.* **36**, 566–574 (1989).
18. Grandy, D. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 9762–9766 (1989).
19. Kobilka, B. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 46–50 (1987).

20. Frielle, T. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7920-7924 (1987).
21. Cotecchia, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7159-7163 (1988).
22. Kobilka, B. K. et al. *Science* **238**, 650-656 (1987).
23. Strader, C. D. et al. *J. biol. Chem.* **263**, 10267-10271 (1988).
24. Venter, J. C., Fraser, C. M., Kerlavage, A. R. & Buck, M. *Biochem. Pharmacol.* **38**, 1197-1208 (1989).
25. Kishimoto, A. et al. *J. biol. Chem.* **260**, 12492-12499 (1985).
26. Bouvier, M. et al. *Nature* **333**, 370-374 (1988).
27. Cullen, B. R. *Meth. Enzym.* **152**, 684-704 (1987).
28. Niznik, H. B. et al. *Molec. Pharmacol.* **34**, 29-36 (1990).
29. Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745-2752 (1987).
30. Brown, B. L., Albano, J. D., Elkins, R. P. & Sgherzi, A. M. *Biochem. J.* **171**, 561-562 (1971).

ACKNOWLEDGEMENTS. We thank L. Robbins for help with cAMP measurements, K. Neve for suggestions, and C. Spear and L. Raymond for help with tissue culture. We thank R. Goodman and M. Forte for manuscript review and J. Tasnady and J. Shiigi for manuscript preparation. This work was supported by the NIH (O.C.).

Human dopamine D₁ receptor encoded by an intronless gene on chromosome 5

Roger K. Sunahara*, Hyman B. Niznik*^{†‡},
David M. Weiner§, Tom M. Stormann§,
Mark R. Brann§, James L. Kennedy¶,
Joel E. Gelernter¶, Richard Rozmahel||, Yili Yang||,
Yedy Israel*#, Philip Seeman*[†] & Brian F. O'Dowd*#

Departments of *Pharmacology and †Psychiatry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

§Laboratory of Molecular Biology, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland 20982, USA

¶Department of Human Genetics and Psychiatry, Yale University, School of Medicine, New Haven, Connecticut 06521, USA

||Department of Genetics, Research Institute, The Hospital for Sick Children, Toronto, Canada

#Addiction Research Foundation, 33 Russell St, Toronto, Ontario M5S2S1, Canada

RECEPTORS for dopamine have been classified into two functional types, D₁ and D₂ (refs 1, 2). They belong to the family of receptors acting through G (or guanine nucleotide-binding) proteins³. D₂ receptors inhibit adenylyl cyclase, but D₁ receptors stimulate adenylyl cyclase and activate cyclic AMP-dependent protein kinases^{4,5}. Dopamine D₁ and D₂ receptors are targets of drug therapy in many psychomotor disorders, including Parkinson's disease and schizophrenia^{6,7}, and may also have a role in drug addiction and alcoholism⁸. D₁ receptors regulate neuron growth and differentiation⁹, influence behaviour and modify dopamine D₂ receptor-mediated events^{10,11}. We report here the cloning of the D₁ receptor gene, which resides on an intronless region on the long arm of chromosome 5, near two other members of the G-linked receptor family. The expressed protein, encoded by 446 amino acids, binds drugs with affinities identical to the native human D₁ receptor. The presence of a D₁ receptor gene restriction fragment length polymorphism will be helpful for future disease linkage studies.

Our goal was to find the full-length DNA sequence from the previously reported¹² partial rat clone of 450 base pairs (bp) (G36), as it was apparent that this clone encoded another G-linked receptor protein. We used G36 to probe a λEMBL 3 SP6-T7 human genomic library (Clontech), and isolated three positive clones with inserts of about 14 kilobases (kb). Sequence and restriction map analysis indicated that one of these clones (RS1) contained sequence similar to the rat complementary DNA clone (G36). Sequencing of overlapping restriction fragments of RS1 (Fig. 1a) revealed a long open reading frame of 1,476 nucleotides (Fig. 1b) encoding a 446-amino-acid protein.

A comparison of the amino-acid sequence of the RS1 with

related receptors reveals that 36 residues are identical, and 72 are conservatively substituted (Fig. 1c), mostly within the seven putative transmembrane domains or in several of the small connecting loops. The percentage of amino acids identical with those in RS1, within transmembrane regions (175 residues) and overall for the entire protein are as follows: β₁-adrenergic receptor, 44% and 33%; β₂-adrenergic receptor, 46% and 30%; platelet α₂-adrenergic receptor, 40% and 24%; dopamine D₂ receptor, 44% and 29%; 5-hydroxytryptamine 5HT_{1A} receptor, 44% and 28%; and muscarinic M1 receptor, 31% and 24%. In common with the human β₂-adrenergic receptor, RS1 contains putative glycosylation sites in both the amino terminus and in the extracellular loop between transmembrane four and transmembrane five (Fig. 1b). Consensus sequences for phosphorylation by cAMP-dependent protein kinase appear twice in the sequence of RS1, and there are three sites for phosphorylation by protein kinase C (Fig. 1b). There are 17 serine and threonine residues in the carboxy terminus of RS1, which may serve as substrates for the enzyme β-adrenergic receptor kinase. The protein has a predicted relative molecular mass of 49,261 (M_r 49.2K) and is similar in electrophoretic mobility (M_r ~ 46K) to the deglycosylated photoaffinity-labelled D₁ receptor, as determined by SDS-PAGE¹³. The structural features that vary appreciably among the receptors aligned in Fig. 1c are the sizes of the third cytoplasmic loop and carboxy-terminal tail. In this context the amino-acid sequence encoded by RS1 more closely resembles the β₂-adrenergic receptor (short third cytoplasmic loop and long carboxy tail) than the D₂-dopamine receptor (long third cytoplasmic loop and short carboxy tail)¹⁴. Although the exact significance of this close structural similarity between the RS1 and β₂-adrenergic receptor is not clear, it may be a consequence of the requirement of these receptors to interact with common cytoplasmic transducing molecules (G proteins)^{15,16}.

To establish that clone RS1 encoded the human D₁ receptor we subcloned a 2.6-kb *NcoI*-*SalI* fragment from RS1 into the multiple cloning site of the eukaryotic expression vector pBC12B (ref. 17) and assessed the ability of the selective dopamine D₁ receptor antagonist [³H]SCH-23390 to bind the membranes of transfected COS-7 cells. COS-7 cells transfected with this construct bound [³H]SCH-23390 with high affinity (dissociation constant, K_d, 630 pM; identical to the K_d for native human caudate tissue^{6,7}) and expressed quantities of receptor protein ranging from 510 to 890 fmol per mg protein (Fig. 2a). Membranes prepared from control cells did not display specific [³H]SCH-23390 binding activity.

Dopaminergic agonists and antagonists inhibited the binding of [³H]SCH-23390 with an appropriate rank order of potency for the D₁ dopamine receptor (Fig. 2b). The selective dopamine D₂ receptor agonist, LY-171555, and antagonist, eticlopride, were virtually without effect on [³H]SCH-23390 binding. The inhibition constant, K_i, values for these compounds correlated well with K_i values obtained in human native membranes (Fig. 2c). Moreover, K_i values of dopaminergic agonists and antagonists for the cloned human D₁ receptor correlate well with both the K_a (association constant) and K_i values of these drugs for dopamine-stimulated adenylyl cyclase activity in dispersed cells of the bovine parathyroid gland¹⁸ (Fig. 2d). Unlike that seen with native membrane preparations, however, agonist versus [³H]SCH-23390 competition curves were uniphasic (Fig. 2b), consistent with the predominance of a non-guanine nucleotide-sensitive low-affinity form of the receptor¹⁸.

Northern blot analysis with the rat cDNA clone, G36, (Fig. 3a) revealed the presence of a 3.8-kb RNA species in RNA extracted from rat striatum and olfactory tubercles, with trace amounts observed in frontal cortex and hypothalamus. No detectable RNA species were observed in rat olfactory bulbs, hippocampus, cerebellum, the anterior and neurointermediate lobes of the pituitary gland, and in rat kidney, liver, lung and heart tissues (data not shown). In the bovine parathyroid gland two RNA species of 4.2 and 3.8 kb were observed (Fig. 3a) and

* To whom correspondence should be addressed.

FIG. 1. *a*, Restriction map of the human dopamine D₁ receptor genomic clone. Relevant endonuclease restriction sites of the 2.9-kb genomic clone; *EcoRI*(E), *NcoI*(Nc), *BglII*(B), *PvuII*(P), *NdeI*(N), *HindIII*(H), *HgaI*(Hg) and *AatII*(A), are indicated. Hatched area, coding region of the receptor with putative transmembrane domains (I-VII) demarcated by solid lines. *b*, Nucleotide sequence of RS1 genomic DNA and deduced amino-acid sequence (single-letter code) of the human dopamine D₁ receptor. The nucleotides are numbered on the right-hand side beginning with the nucleotide triplet coding for the putative initiation methionine and ending at the termination codon. The double underline denotes the small open reading frame in the 5' untranslated region. The position of the transmembrane (TM) regions are indicated by shaded areas. Also shown are, postulated *N*-linked glycosylation sites (▲), and putative protein kinase A (□) and protein kinase C (■) phosphorylation sites. Cys 348 is shown boxed, an equivalent cysteine residue in β -adrenergic receptor is modified by covalent attachment of palmitic acid²⁴. *c*, Alignment of receptor primary structures. Deduced protein sequences of the human dopamine D₁ receptor (D1), the human β_1 adrenergic receptor (BETA-1), the human β_2 adrenergic receptor (BETA-2), the human α_2 adrenergic receptor (ALPHA-2C10), the long form of the human dopamine D₂ receptor (D2), the human serotonin (5-hydroxy tryptamine) 1A receptor

(S1A), and the human muscarinic M1 receptor (M1) were aligned to optimize homology. The approximate positions of the putative transmembrane regions are indicated by the horizontal lines above the sequences and the conserved amino acids are boxed in colour (pink for those conserved in the transmembrane region, yellow for those conserved outside the transmembrane regions).

METHODS. Nick-translated (Amersham) G36, a 500-bp DNA fragment which encodes part of the rat dopamine D₁ receptor, was used to probe a λ EMBL 3 human genomic library (Clontech). Duplicate filters (Hybond, Amersham) were hybridized at 42 °C in medium containing ³²P-labelled G36 (2 × 10⁶ c.p.m. ml⁻¹), 0.1% glycine, 3 × SSC buffer, 100 mM sodium phosphate, pH 6.8, 50% deionized formamide, 1X Denhardt's solution, 10% dextran sulphate and 150 μ g ml⁻¹ denatured and sheared salmon sperm DNA (Sigma). Membranes were washed at 20 °C for 1 h with 2 × SSC and 0.1% SDS, followed by a 51 °C wash for 1 h with 0.2 × SSC and 0.2% SDS. Of 6 × 10⁵ recombinants screened, two identical clones of about 14 kb were plaque-purified, using identical hybridization and washing conditions and subjected to restriction mapping and Southern blot analysis. A 4-kb insert in pSP73 (Promega), containing the full coding region of the human dopamine D₁ receptor flanked by 1.2 kb of 5' and 1.6 kb of 3' untranslated regions,

was constructed as follows: a *HindIII* and *BglII* restriction endonuclease digestion of the 14-kb genomic insert yielded two fragments of 2 and 2.5 kb in length, respectively. Sequence analysis revealed that the fragments contained 500 bp of overlapping sequence which was subsequently excised and the fragments ligated with DNA T4 ligase (Boehringer-Mannheim) and subcloned into pSP73 (Promega). For sequencing, overlapping restriction fragments of the 4-kb insert were subcloned into pSP73 and sequenced in both directions using Sequenase v2.0 (USB) and oligonucleotides complementary to SP6 and T7 promoter regions (Promega) and specific synthetic oligonucleotides (Biotechnology Service Centre, Hospital for Sick Children, Toronto) as internal primers.

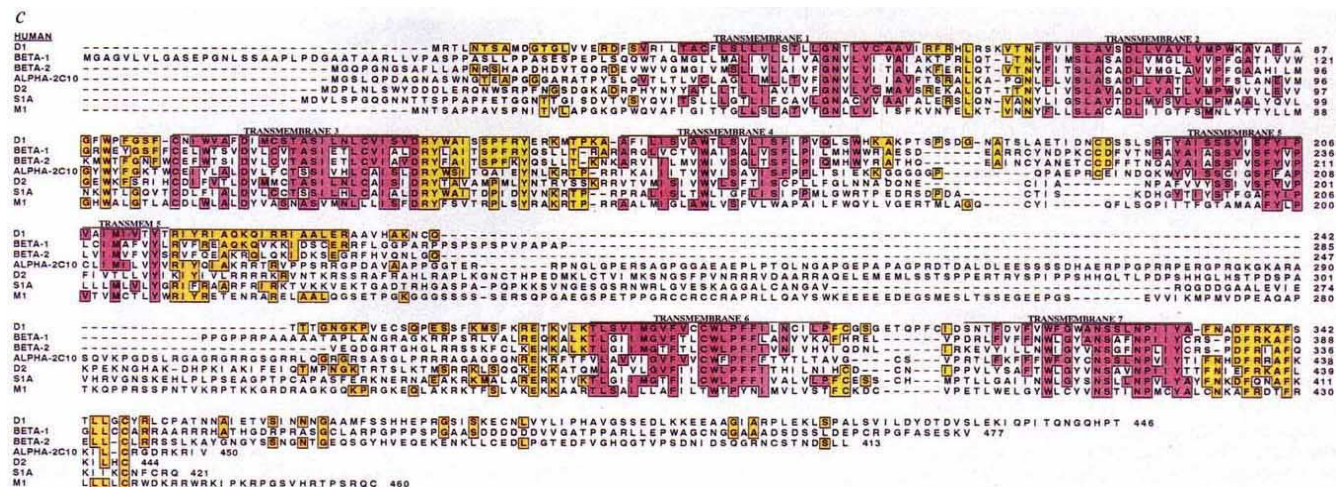
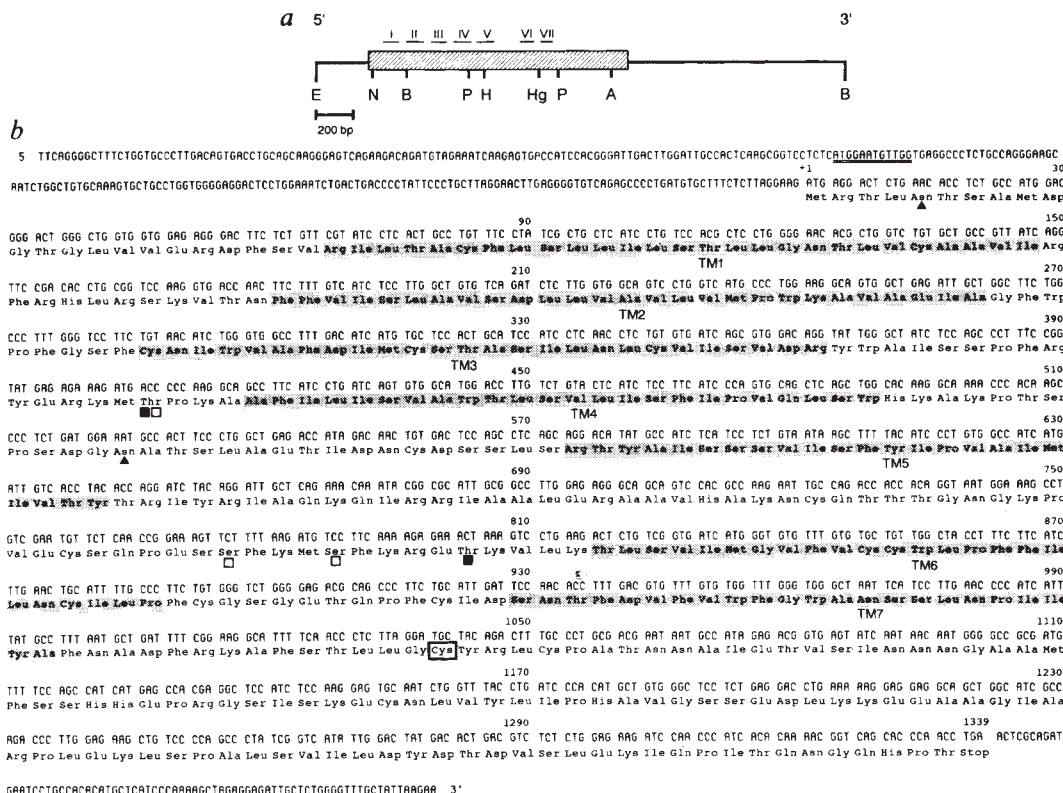


FIG. 2 Binding of [3 H]SCH-23390 to transformed COS-7 cell membranes. *a*, Saturation isotherms of the specific binding of [3 H]SCH-23390 to membranes from COS-7 cells expressing the human dopamine D₁ receptor. The results shown are representative of three independent experiments each conducted in duplicate. Inset, Scatchard plot of the same data. The estimated maximal binding, B_{max} (that is, receptor density; 889 fmol per mg protein), and K_d (656 ± 83 pM) values were obtained by LIGAND, as previously described¹⁸. *b*, Pharmacological specificity of [3 H]SCH-23390 binding to COS-7 cells. Representative curves are shown for the concentration-dependent inhibition of [3 H]SCH-23390 binding by various dopaminergic agonists and antagonists. Data were analysed by LIGAND and the results shown are the means of duplicate determinations and are representative of two experiments. *c*, Line of identity of estimated K_i values of dopaminergic ligands for [3 H]SCH-23390 binding to transfected COS-7 cells and human brain membranes. *d*, Line of identity of K_i values for cAMP accumulation in bovine parathyroid gland¹⁸ and [3 H]SCH-23390 binding to COS-7 cells of various dopaminergic agonists and antagonists.

METHODS. A construct containing a 2.6-kb *Nco*I–*Sal*I fragment was made as follows: pBC12B (ref. 17) containing the β_2 -adrenergic receptor was digested with *Nco*I and *Sal*I, the coding and 3' untranslated region of the β_2 -adrenergic receptor excised and replaced with the 2.6-kb *Nco*I–*Sal*I human dopamine D₁ fragment with the appropriate 5'–3' orientation. Caesium chloride-purified vector DNA was introduced into COS-7 cells by calcium phosphate-mediated transfection as described¹⁹. Cells were collected and homogenized (teflon pestle) in 50 mM Tris-HCl (pH 7.4 at 4 °C) buffer containing 5 mM EDTA, 1.5 mM CaCl₂, 5 mM MgCl₂, 5 mM KCl and 120 mM NaCl. Homogenates were centrifuged for 15 min at 39,000g, and the resulting pellets resuspended in buffer at 100–150 $\mu\text{g ml}^{-1}$. For saturation experiments, 0.25 ml aliquots of tissue homogenate were incubated in duplicate with increasing concentrations of [3 H]SCH-23390 (73.1 Ci mmol⁻¹; 50–6,000 pM final concentration) for 120 min at 22 °C in a total volume of 1 ml. For competition binding experiments, assays

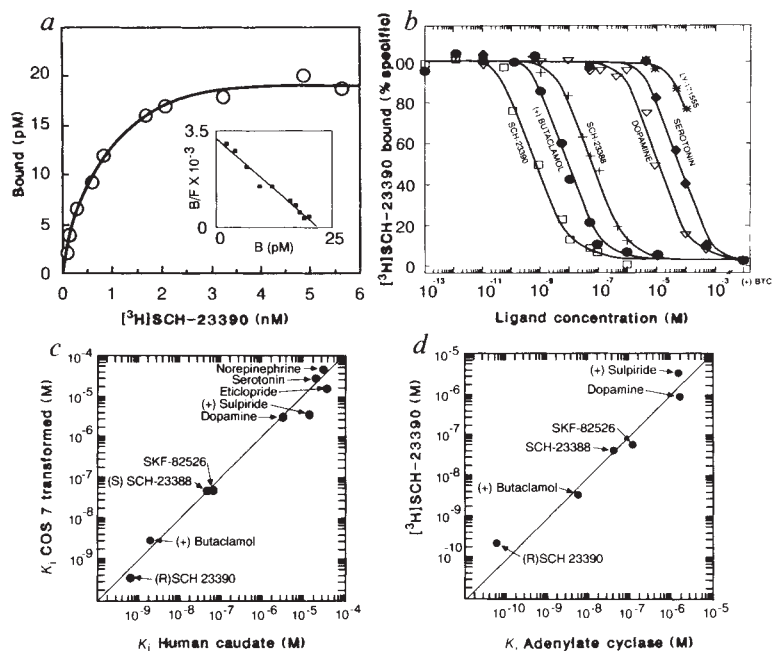
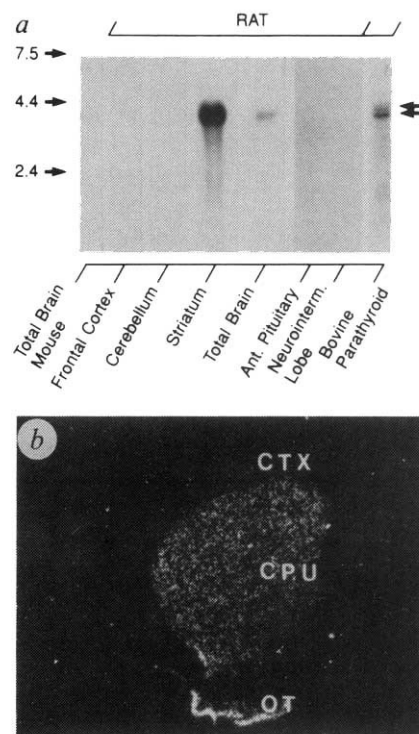


FIG. 3 *a*, Northern blot analysis of the tissue distribution of the dopamine D₁ receptor mRNA. Each lane contains 20 μg total RNA from different brain regions and peripheral tissues. Size markers are shown on the left in kb (BRL). *b*, Localization of the dopamine D₁ receptor mRNA in rat forebrain by *in situ* hybridization histochemistry. Regions that express dopamine D₁ receptor mRNA, including the caudate-putamen, nucleus accumbens, and olfactory tubercle, appear white in the photomicrograph after a five-day exposure. Cortex (CTX), caudate-putamen (CPU) and olfactory tubercle (OT) are indicated.

METHODS. Bovine parathyroid glands and various brain regions from male Wistar rats were dissected and RNA isolated as described²⁵. Tissues (~1 g each) were homogenized in a solution containing 4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0, 0.5% sarcosyl, 0.1 M 2-mercaptoethanol (solution A). Sequentially, 0.1 ml of 2 M sodium acetate, pH 4.0, 1 ml of phenol and 0.2 ml chloroform were added, with mixing after each addition, and placed on ice for 15 min. The solutions were centrifuged at 8,000g for 25 min at 4 °C and the aqueous phase mixed with isopropanol (1 ml) and placed at –20 °C for one hour. The RNA was pelleted by centrifugation at 8,000g for 20 min and the pellets resuspended in 0.3 ml of solution A and precipitated with 1 vol of isopropanol at 20 °C for 1 h. Following centrifugation, RNA was washed with 70% ethanol and dried, dissolved in water and subjected to electrophoresis through 6% formaldehyde and 1% agarose gels and blotted. Nitrocellulose membranes were prehybridized at 42 °C (overnight) in medium containing 0.1% glycine, 3 × SSC, 100 mM sodium phosphate, pH 6.8, 50% deionized formamide, 1 × Denhardt's, 0.5% SDS and 150 $\mu\text{g ml}^{-1}$ denatured and sheared salmon sperm DNA (Sigma). Subsequent hybridization was performed at 42 °C (overnight) in medium (as described above) except with the addition of ³²P-labelled G36 (2×10^6 c.p.m. ml⁻¹) and 10% dextran sulphate. Membranes were washed at 20 °C for 1 h with 2 × SSC and 0.1% SDS followed by a 60 °C wash for 1 h with 0.2 × SSC and 0.2% SDS. Blots were exposed to Kodak XAR autoradiographic film with one intensifying screen at –70 °C for 24–72 h. *b*, *In situ* hybridization was as previously described²⁶. Briefly, 12- μm flash-frozen rat brain sections were cut onto gelatin-coated slides at –20 °C. A mixture of three oligonucleotide probes complementary to sequences in G36 D_{1A} (5'-TAAAGGTTTACACTGGGCCCATTCGACGCGTTCCCGTTACCTG-3') D_{1B} (5'-GGACTGCTGCCCTCTCCAAGGCTGAGATGCCCGGATTGCTTCTGGG-3') and D_{1C} (5'-TGTCACACTTGTATCCTCGGTGCTCCAGGAGGTAATAATGCCAT-3') were end-labelled with [³⁵S]dATP (NEN) using terminal deoxynucleotidyl

were initiated by the addition of 0.5 ml of membrane and incubated in duplicate with the indicated concentrations of competing ligands (10^{-14} – 10^{-3} M) and [3 H]SCH-23390 (500–600 pM) in the presence of 120 mM sodium chloride for 120 min at 22 °C. Assays were terminated by rapid filtration through a Titertek cell harvester and filters subsequently monitored for tritium as described¹⁸. For all experiments, specific [3 H]SCH-23390 binding was defined as that inhibited by 1 μM (+)-butaclamol. Both saturation and competition binding data were analysed by the nonlinear least-square curve-fitting program LIGAND run on a Digital Micro-PDP-11.



transferase (BRL). Sections were fixed in 4% formaldehyde for 5 min, blocked with 0.25% acetic anhydride in 0.1 M triethanolamine with 0.9% NaCl for 10 min, dehydrated and delipidated. Oligonucleotide probes (10^7 d.p.m.) were incubated in hybridization medium for 24 h at 37 °C, washed four times for 45 min in 1 × SSC at 57 °C and then washed twice for 45 min in the same buffer at 20 °C. Fixed sections were placed against autoradiographic film for 5 days or dipped in NTB3 nuclear tract emulsion (Kodak mixed 1:1 with H₂O) and exposed for 14 days.

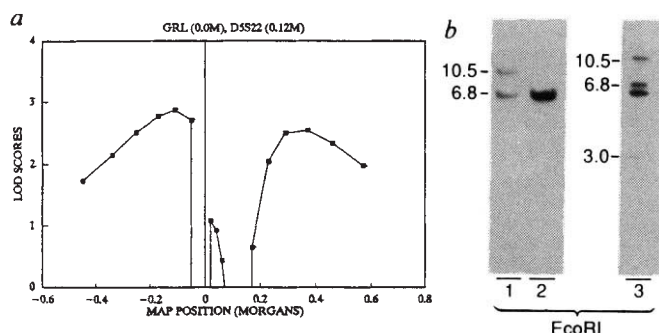


FIG. 4 *a*, Multipoint analysis of the dopamine D_1 receptor locus. The dopamine receptor D_1 locus, recognized by an *EcoRI* polymorphism, using probe G36, *EcoRI*-digested DNA from unrelated individuals identified a simple two-allele RFLP; allelic frequencies for the A1 (10.5 kb) allele, 0.28, and the A2 (6.8 kb) allele, 0.72—was moved across a fixed map of GRL placed arbitrarily at 0.00M, and D5S22 placed at 0.12M. Multipoint analyses employed the LINKAGE package²⁷. Results are the sum of lod scores obtained with two families, TSCANB and TSCANC. The maxima on opposite sides of the GRL–D5S22 interval are not significantly different and therefore the dopamine receptor D_1 locus cannot be ordered by these results. *b*, Human DNA (3 μ g) digested with *EcoRI* and subjected to electrophoresis on a 1% agarose gel, Southern blotted²¹ and probed with G36 (lane 1 and 2) and the D_1 receptor 2.5-kb fragment (lane 3). These probes recognize the two-allele RFLP (lane 1 and 3 heterozygous, lane 2 homozygous).

probably represent transcripts from different polyadenylation sites.

To determine the expression pattern of the D_1 receptor in rat, *in situ* hybridization histochemistry was performed on brain slices. Three oligonucleotides based on the sequence of the rat G-36 clone¹² were synthesized. The probes were complementary to regions of the molecule that displayed the lowest degree of sequence homology to other members of this gene family, notably in the extracellular domain between TM4 and TM5 and the large third cytoplasmic loop. Each of the probes were used individually for *in situ* analysis on rat forebrain where they exhibited an identical pattern of labelling. A mixture of the three probes was used to increase the sensitivity of detection. The strongest signal was seen in the striatum (Fig. 3*b*), including the caudate–putamen, nucleus accumbens, and olfactory tubercle; 50% of the medium-sized striatum neurons expressed this messenger RNA. Moderate signals were observed in the frontal cortex, habenula, entorhinal cortex, nuclei in the amygdala and hypothalamus, and thalamus, with little or no hybridization seen in other brain regions including the substantia nigra. The distribution of expression of this receptor correlates with the dopaminergic innervation, and with the distribution of D_1 receptor sites^{19,20}.

Southern blot hybridization to a hybrid cell panel²¹ revealed chromosome 5 to be the only human chromosome that segregated perfectly with the D_1 dopamine receptor gene; all the others were discordant by two or more hybrids (data not shown). To corroborate and refine the hybrid cell panel localization, genetic linkage analyses were performed by typing the D_1 dopamine receptor restriction fragment length polymorphism (RFLP) in two branches of a large reference kindred at Yale. The rat G36 clone, the human genomic D_1 receptor fragment (2.5 kb), and the pHD₁-Gem clone from Civelli *et al.* (personal communication) recognized the same polymorphism. Two markers in distal 5q, namely GRL and D5S22, showed positive lod scores at varying recombination distances from the D_1 dopamine receptor locus. Obligate crossovers were revealed with each of the loci. Negative pairwise lod scores with 17 markers ruled out linkage of the D_1 dopamine receptor locus to a wide range of other locations on chromosome 5. A multipoint analysis using both GRL and D5S22 (GRL is 12 cM from D5S22 based on the data of Giuffra *et al.*²²) is shown in Fig. 4. Given the above evidence for the presence of the D_1 dopamine receptor

locus on chromosome 5, the peak lod score of 2.89 confirms that the D_1 receptor locus maps to this chromosome and is located near GRL and D5S22 in the 5q31–34 region near the structurally homologous genes for β_2 -adrenergic receptor and α_1 -adrenergic receptor²³. □

Received 30 April; accepted 18 June 1990.

1. Kebabian, J. W. & Calne, D. B. *Nature* **277**, 93–96 (1979).
2. Seeman, P. *Pharmacol. Rev.* **32**, 229–313 (1980).
3. Freissmuth, M., Casey, P. J. & Gilman, A. G. *FASEB J.* **3**, 32125–32131 (1989).
4. Niznik, H. B. *Molec. cell. Endocrinol.* **54**, 1–22 (1987).
5. Hemmings, H. C., Walaas, S. I., Olumet, C. C. & Greengard, P. *Trends Neurosci.* **10**, 77–82 (1987).
6. Seeman, P. *et al.* *Neuropsychopharmacol.* **1**, 5–15 (1987).
7. Seeman, P. *Synapse* **1**, 333–352 (1987).
8. Blum, K. *et al.* *JAMA* **263**, 2055–2060 (1990).
9. Lankford, K. L., Demello, F. G. & Klein, W. L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2839–2843 (1988).
10. Clark, D. & White, F. J. *Synapse* **1**, 347–388 (1987).
11. Seeman, P., Niznik, H. B., Guan, H. C., Booth, G. & Ulpian, C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 10156–10160 (1989).
12. O'Dowd, B. F. *et al.* *FEBS Lett.* **262**, 8–12 (1990).
13. Jarvie, K. R., Booth, G., Brown, E. M. & Niznik, H. B. *Molec. Pharmacol.* **34**, 566–574 (1989).
14. Bunzow, J. R. *et al.* *Nature* **336**, 783–787 (1988).
15. O'Dowd, B. F., Lefkowitz, R. J. L. & Caron, M. G. A. *Rev. Neurosci.* **12**, 67–83 (1989).
16. O'Dowd, B. F. *et al.* *J. Biol. Chem.* **263**, 15985–15992 (1988).
17. Cullen, B. *Meth. Enzym.* **152**, 684–704 (1987).
18. Niznik, H. B., Fogel, E. L., Chen, C. J., Brown, E. M. & Seeman, P. *Molec. Pharmacol.* **34**, 29–36 (1988).
19. Bjorklund, A. & Lindvall, O. in *Handbook of Chemical Neuroanatomy* (eds Bjorklund, A. & Hokfelt, T.) vol. 2 (A) 55–122 (Elsevier, Amsterdam, 1984).
20. Savasta, M., Dubois, A. & Scatton, G. *Brain Res.* **375**, 291–301 (1986).
21. O'Dowd, B. F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 1184–1188 (1985).
22. Giuffra, L. A. *et al.* *Cytogen. Cell. Gen.* **49**, 313–314 (1988).
23. Yang-Feng, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 1516–1520 (1990).
24. O'Dowd, B. F. *et al.* *J. Biol. Chem.* **264**, 7564–7569 (1989).
25. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
26. Weiner, D. M. & Brann, M. R. *FEBS Lett.* **253**, 207–213 (1989).
27. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Am. J. hum. Genet.* **37**, 482–498 (1985).

ACKNOWLEDGEMENTS. We thank T. Nguyen, A. Tirpak, C. Ulpian, M. Kertesz, I. Novotni, S. Bisighini, G. Schofield for technical assistance, and Lap-Chee Tsui for the chromosomal mapping experiment (blot from R. G. Lafreniere and H. F. Willard, Stanford University). For their help, we thank D. Mahuran and C. Brown (COS-7 expression studies), L. Giuffra and K. Kidd (linkage studies), and M. Hnatowich (sequence alignment). Finally, we thank S. George, R. Dunn and all our colleagues for discussion. This work was supported in part by The Addiction Research Foundation of Ontario (ARF), the National Alliance for Research on Schizophrenia and Depression (NARSAD), the Ontario Mental Health Foundation, by an ARF predoctoral fellowship to R.K.S., National Institute of Mental Health, the Scottish Rite Schizophrenia Research Program and by the Veteran's Administration National Center for PTSD. Both H.B.N. and J.L.K. are recipients of NARSAD Young Investigator Awards and H.B.N. is a Career Scientist of the Ontario Ministry of Health. D.M.W. is a recipient of a NIH–Howard Hughes Medical Institute Research Scholars Program.

Photoreceptor degeneration in inherited retinal dystrophy delayed by basic fibroblast growth factor

Ella G. Faktorovich, Roy H. Steinberg, Douglas Yasumura, Michael T. Matthes & Matthew M. LaVail*

Departments of Anatomy, Physiology and Ophthalmology, and Beckman Vision Center, University of California at San Francisco, San Francisco, California 94143-0730, USA

NUMEROUS inherited retinal degenerations exist in animals and humans, in which photoreceptors inexplicably degenerate and disappear. In RCS rats with inherited retinal dystrophy, the mutant gene is expressed in the retinal pigment epithelial (RPE) cell, and leads to the loss of photoreceptor cells¹. Photoreceptors can be rescued from degeneration if they are juxtaposed to wild-type RPE cells in experimental chimeras¹ or by the transplantation of RPE cells from normal rats^{2,3}. In both cases, the rescue effect extends beyond the immediate boundaries of the normal RPE cells, suggesting trophic action of a diffusible factor(s) from the normal RPE cells. We considered that the fibroblast growth factors, aFGF and bFGF, might have such a trophic role as they are found in the

* To whom correspondence should be addressed.

retina⁴⁻⁸ and RPE cells^{9,10}; bFGF acts as a neurotrophic agent after axonal injury in several regions of the central nervous system¹¹⁻¹³, and bFGF induces retinal regeneration from developing RPE cells¹⁴. Here we report that subretinal injection of bFGF results in extensive rescue of photoreceptors in RCS rats for at least two months after the injection, and that intravitreal injection of bFGF results in even more widespread rescue, across almost the entire retina. The findings demonstrate for the first time that bFGF can act as a survival-promoting neurotrophic factor in a hereditary neuronal degeneration of the central nervous system.

We injected aFGF or bFGF into either the subretinal (interphotoreceptor) space or vitreous of RCS rats at postnatal day (P) 23, when photoreceptor degeneration had just begun¹⁵ (Fig. 1a). For buffer control experiments we injected an equal volume of phosphate-buffered saline (PBS), and for surgical controls we inserted a dry needle either into the subretinal space or through the retina into the vitreous. We then examined the retina by light microscopy at either one month (Fig. 1b, c) or two months (Fig. 2) post-injection.

After one month (P53), uninjected control retinæ showed the expected loss of most photoreceptor cells¹⁵, with a reduction of the outer nuclear layer (ONL) to one to two rows of photoreceptor nuclei, many of which were pyknotic, and the loss of virtually all photoreceptor inner segments (Fig. 1b). The outer segment region had become a debris zone of disorganized and degenerating membranes owing to a defect in RCS rats in the normal RPE function of phagocytosing outer segment membranes^{16,17}. When bFGF was injected into the subretinal space, retinæ at P53 showed a remarkable preservation of many photoreceptors in the superior hemisphere of the eye in which the injection had been made (Fig. 1c). In regions of optimal rescue (Fig. 1c), the retina appeared as intact as at the time of injection (compare with Fig. 1a). Despite the preservation of photoreceptor cells, bFGF clearly had not reversed the RPE phagocytosis defect, on the basis of the absence of phagosomes in the RPE cells and the presence of a substantial debris zone (Fig. 1c).

After two months (P83), the ONL of uninjected control retinæ showed no more than one row of photoreceptor nuclei¹⁵. Following subretinal bFGF injection, however, the ONL in a few regions was still almost as thick as at one month, but in most retinæ maximal ONL thickness had been reduced from eight to ten rows of nuclei to five to seven rows (Fig. 2a). In these areas of photoreceptor rescue there were relatively few pyknotic nuclei, as at one month post-injection, and photoreceptor inner segments were usually evident wherever the ONL was two to three or more nuclei in thickness. Examination of serial sections showed that the rescue often extended throughout the superior hemisphere of the eye, but usually not into the inferior hemisphere (Figs 3a and 4). When bFGF was injected intravitreally, the extent of photoreceptor rescue was far greater, encompassing almost the entire retina (Figs 3 and 4). In general, the overall ONL thickness in eyes intravitreally injected with bFGF was less than the maximal ONL thickness observed in eyes subretinally injected (Fig. 4b, c).

The injection of aFGF or PBS either subretinally or intravitreally also produced photoreceptor rescue, as did the surgical controls, confirming a recent report that PBS injected subretinally rescues photoreceptors¹⁸. But in most cases this rescue was much more limited in extent than rescue produced by bFGF (Figs 2b, 3 and 4), because it was usually localized to the site of injection or surgery as determined in serial sections. In the region of optimal rescue, however, aFGF, PBS and surgical controls sometimes produced the same ONL thickness as bFGF and vitreal bFGF (for example, see height of peaks in Fig. 3a). Regardless of the procedure that produced the rescue, photoreceptor inner segments were present whenever the ONL was two to three or more rows of nuclei in thickness. Thus, the limited region of optimal rescue in the controls seemed qualitatively similar to that produced by bFGF. The principal difference was the much greater extent of rescue produced by

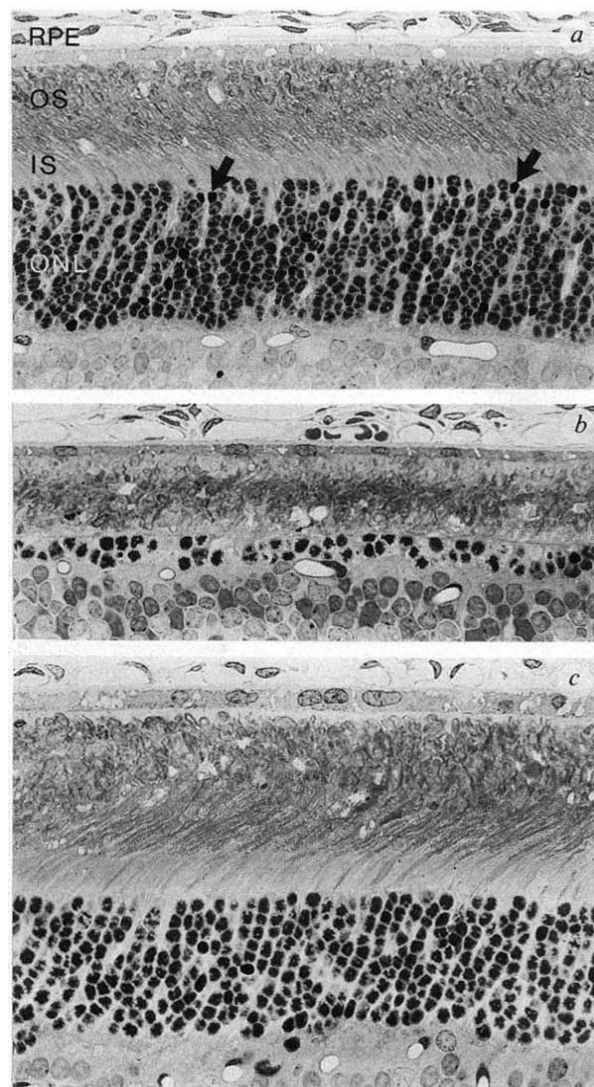
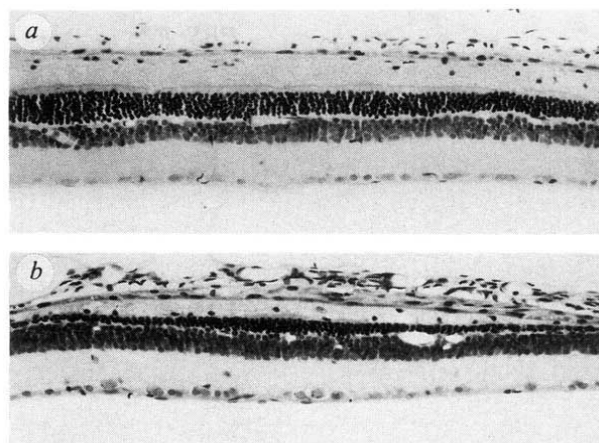


FIG. 1 Plastic-embedded sections of RCS rat retinæ. a, P23, at the time of injections; cell death had just begun and some photoreceptor nuclei in the ONL were pyknotic (arrows). Photoreceptor inner segments (IS) and outer segments (OS) are present, as well as some accumulated outer segment debris membranes near the apical surface of the RPE. b, P53, an uninjected control animal; the ONL is reduced to one to two rows in thickness, no photoreceptor inner or outer segments are present and the outer segment zone is composed only of debris membranes. c, P53, a region of optimal photoreceptor rescue 1 month after subretinal bFGF injection; photoreceptor inner and outer segments are present and the retina generally appears unchanged from the time of injection. The retina in the opposite hemisphere of the same eye as in c was fully degenerated and indistinguishable from that shown in b. Toluidine blue stain. Scale bar, 10 μ m.

METHODS. Brain-derived bovine aFGF and recombinant human bFGF, both affinity-purified, were provided by D. Gospodarowicz. Specific activity of aFGF was 1 ng ml⁻¹ protein and bFGF was 50–80 pg ml⁻¹ protein for 50% effective dose (ED₅₀) in 3T3 fibroblasts. Subretinal injections were made transclerally in the superior hemispheres of the eyes of anaesthetized rats²⁸, as were the intravitreal injections, which avoided surgery. The volume of injection was 1 μ l in all cases. The aFGF concentration was 500 ng μ l⁻¹ and the bFGF was 500 or 820 ng μ l⁻¹; both concentrations of bFGF gave identical results so the data for the two were pooled. The rats were killed by overdose of carbon dioxide followed by vascular perfusion of mixed aldehydes, and the eyes were embedded in epoxy resin for 1- μ m-thick sections or in polyester wax for 10- μ m-thick serial sections by methods previously described¹⁵. The eyes were sectioned near or parallel to the vertical meridian to allow comparison of the injected superior hemisphere with the noninjected inferior hemisphere.

FIG. 2 Polyester wax-embedded sections of RCS rat retinas at P83, 2 months after *a*, subretinal injection of bFGF, showing substantial photoreceptor rescue; and *b*, the surgical control of a needle in the subretinal space, showing rescue limited to the region appearing in the micrograph. The limited rescue in *b* was typical of PBS, aFGF and the other surgical controls. Scale bar, 25 μm .



both subretinal and vitreal bFGF. The rescue produced by aFGF often appeared slightly more extensive than that produced by PBS or surgical controls, but it did not differ statistically from them in quantitative measurements. Thus we cannot attribute a specific rescue effect to aFGF.

The explanation for rescue produced by PBS, aFGF and the surgical controls may be the release of endogenous bFGF when RPE cells or other cells are damaged during injection. A similar release of bFGF has been observed following mechanical damage to vascular endothelial cells¹⁹ or focal brain injury²⁰. This idea is consistent with our finding that the extent of rescue is generally proportional to the degree of injury during the

injection. Mechanisms other than the release of bFGF from damaged cells may be operating. For instance, rescue was significantly more extensive in several cases where excessive bleeding occurred during surgery (Fig. 3*b*), suggesting a possible role for various blood elements (for example, monocyte-derived macrophages) that contain and release bFGF^{21,22}.

Evidence has accumulated for the existence of specific neurotrophic factors that regulate growth and survival of photoreceptors^{23,24}, and RPE produces several molecules that are crucial for photoreceptor growth and survival²⁵. The presence of bFGF in the RPE^{9,10} and retina⁴⁻⁸ and the rescue effect shown here in RCS rats suggest a possible neurotrophic role in

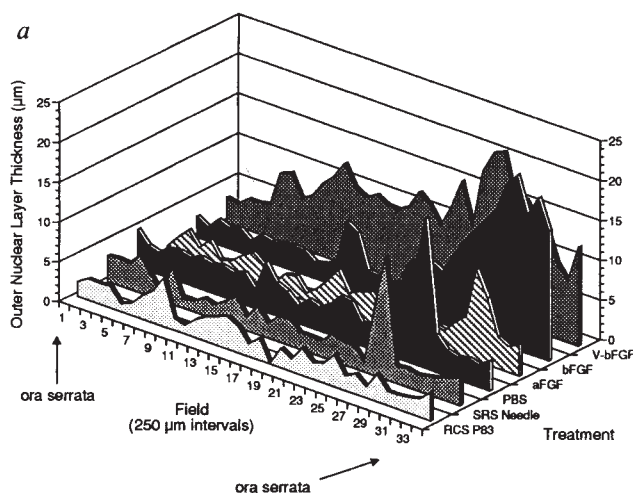
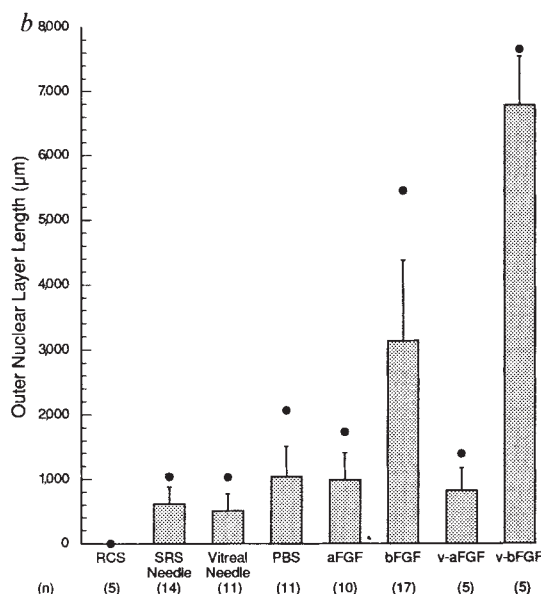


FIG. 3 *a*, Thickness of the ONL in representative sections from RCS rat retinas at P83 where, at P23, bFGF had been injected either subretinally (bFGF) or intravitreally (v-bFGF), aFGF or PBS had been injected subretinally, or a dry needle had been inserted into the subretinal space (SRS needle) as a surgical control. For comparison, a section is shown from an uninjected RCS rat at P83. These sections were taken near the vertical meridian of the eye and illustrate the ONL from the ora serrata in the inferior hemisphere (left) through the centre of the eye (middle) to the ora serrata in the superior hemisphere (right). The eyes receiving aFGF, PBS or sham surgical control showed only a single, restricted peak of photoreceptor rescue, whereas the eye receiving bFGF subretinally showed a broad area of rescue extending almost throughout the superior hemisphere. With vitreal bFGF, photoreceptor rescue extended almost throughout both hemispheres of the retina.

METHODS. ONL thickness was measured with Bioquant morphometry software using a digitizing tablet and a camera lucida. Three measurements were taken at 75- μm intervals within each 250- μm length of retina, and the means of the three measurements were plotted for each 250- μm length of retina. *b*, Maximal lengths of rescued ONL in P83 RCS rats that had received various treatments at P23. Serial sections were examined in each of the eyes (number in parentheses) and the single section with the greatest length of rescue was measured. Each bar represents the mean ($\pm$ s.d.) ONL



length of these sections and the highest value for each group is indicated by $\bullet$. The length of the entire section from one ora serrata to the other was $\sim 8,000 \mu\text{m}$. PBS, aFGF and bFGF are injections into the subretinal space (SRS); v-aFGF and v-bFGF are intravitreal injections; and SRS needle and vitreal needle are surgical controls. A graph of the areas of rescued ONL (not illustrated) measured on the same sections was virtually identical to this graph. Six control eyes (two SRS needle, two vitreal needle and two PBS) were excluded because of extensive intraocular bleeding at the time of surgery or injection. These eyes typically showed large macrophages with brown intracytoplasmic pigment, rosettes of rescued photoreceptor cells, and of most interest, a rescue that was significantly more extensive than in the other controls or aFGF, and which was $\sim 75\%$ the extent of rescue produced by subretinal injection of bFGF. In addition, epiretinal membranes formed in eight eyes (three vitreal needle, one v-aFGF and four v-bFGF) where in our earliest experiments the lens had been damaged or bleeding had occurred during intravitreal injections. This resulted in extensive photoreceptor rescue across most of the retina. These eyes also were not quantified.

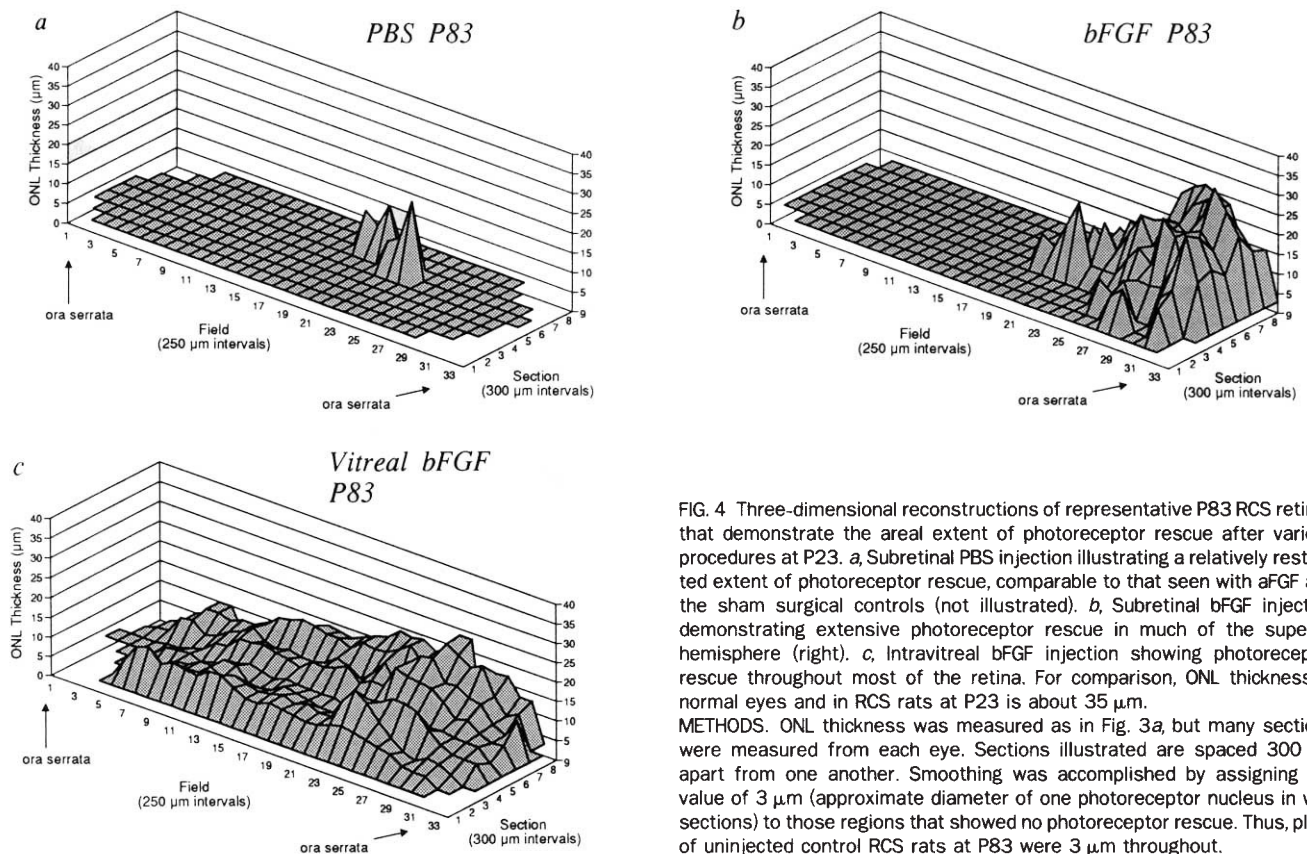


FIG. 4 Three-dimensional reconstructions of representative P83 RCS retinas that demonstrate the areal extent of photoreceptor rescue after various procedures at P23. *a*, Subretinal PBS injection illustrating a relatively restricted extent of photoreceptor rescue, comparable to that seen with aFGF and the sham surgical controls (not illustrated). *b*, Subretinal bFGF injection demonstrating extensive photoreceptor rescue in much of the superior hemisphere (right). *c*, Intravitreal bFGF injection showing photoreceptor rescue throughout most of the retina. For comparison, ONL thickness in normal eyes and in RCS rats at P23 is about 35 μm .

METHODS. ONL thickness was measured as in Fig. 3*a*, but many sections were measured from each eye. Sections illustrated are spaced 300 μm apart from one another. Smoothing was accomplished by assigning the value of 3 μm (approximate diameter of one photoreceptor nucleus in wax sections) to those regions that showed no photoreceptor rescue. Thus, plots of uninjected control RCS rats at P83 were 3 μm throughout.

normal photoreceptor-RPE cell interactions, but this remains to be shown experimentally. Indeed, given the demonstrated interactions between bFGF and other factors, the complexity of photoreceptor-RPE cell interactions, and the possibility that other cells such as macrophages or Müller cells might be involved, it is not certain that the bFGF in the present study acted directly on photoreceptor cells.

The rescue of photoreceptors by bFGF in the RCS rat also might implicate this heparin-binding molecule in the mechanism of this disease. Possibilities include abnormalities in the structure, synthesis, release or delivery of bFGF, or abnormalities in bFGF receptors. Alternatively, bFGF and its receptors may be normal and other conditions may circumvent a normal bFGF role. For example, as bFGF receptors are present on rod outer segments⁸, accumulated outer segment membranes in the debris zone may compete with normal outer segments for available bFGF. Moreover, the interphotoreceptor matrix that accumulates excessively around photoreceptor inner and outer segments in RCS rats²⁶ may, because of its polyanionic properties, inhibit bFGF binding to outer segments²⁷.

What of bFGF as a possible therapeutic agent? On the one hand, bFGF also delays cell loss in light-induced photoreceptor degeneration in the rat (our unpublished observations), so the effect shown here is not restricted to RCS rats and may prove applicable to other retinal degenerations. On the other hand, the functional capacity of rescued photoreceptors, efficacy in other forms of retinal degeneration, longevity of effect, and the relative efficiency of other growth factors or related molecules are not known. For example, it cannot be assumed that a single injection would have a permanent effect, at least in RCS rats, where bFGF does not reverse the genetic defect in outer segment phagocytosis. Indeed, in the few retinas we have examined three months after intravitreal bFGF injection, the ONL was reduced significantly from that at two months, although photoreceptor inner and outer segments were still present. More important are potentially harmful side effects, given the widely known

mitogenic and angiogenic properties of bFGF. In fact, the retinas with intravitreal bFGF injections consistently showed numerous invading macrophages in the inner retina whose origin and role in the rescue effect must be explained. We are presently addressing these and other issues. □

Received 22 May; accepted 22 June 1990.

- Mullen, R. J. & LaVail, M. M. *Science* **192**, 799–801 (1976).
- Li, L. & Turner, J. E. *Exp. Eye Res.* **47**, 911–917 (1988).
- Lopez, R. et al. *Invest. Ophthalmol. Vis. Sci.* **30**, 586–588 (1989).
- D'Amore, P. A. & Klagsbrun, M. *J. Cell Biol.* **99**, 1545–1549 (1984).
- Baird, A., Esch, F., Gospodarowicz, D. & Guillemin, R. *Biochemistry* **24**, 7855–7860 (1985).
- Mascarelli, F., Raulais, D., Counis, M. F. & Courtois, Y. *Biochem. biophys. Res. Commun.* **146**, 478–486 (1987).
- Plouët, J., Mascarelli, F., Loret, M. D., Faure, J. P. & Courtois, Y. *EMBO J.* **7**, 373–376 (1988).
- Plouët, J., in *Molecular Biology of the Eye: Genes, Vision and Ocular Disease* (eds Piatigorsky, J., Tashimichi, S. & Zelenka, P. S.) 83–92 (Liss, New York, 1988).
- Schweigerer, L., Malerstein, B., Neufeld, G. & Gospodarowicz, D. *Biochem. biophys. Res. Commun.* **143**, 934–940 (1987).
- Sternfeld, M. D., Robertson, J. E., Shipley, G. D., Tsai, J. & Rosenbaum, J. T. *Curr. Eye Res.* **8**, 1029–1037 (1989).
- Sievers, J., Hausmann, B., Unsicker, K. & Berry, M. *Neurosci. Lett.* **76**, 157–162 (1987).
- Anderson, K. J., Dam, D., Lee, S. & Cotman, C. W. *Nature* **332**, 360–361 (1988).
- Otto, D., Frotscher, M. & Unsicker, K. *J. Neurosci. Res.* **22**, 83–91 (1989).
- Park, C. M. & Hollenberg, M. J. *Dev. Biol.* **134**, 201–205 (1989).
- LaVail, M. M. & Battelle, B. A. *Exp. Eye Res.* **21**, 167–192 (1975).
- Herron, W. L., Riegel, B. W., Myers, O. E. & Rubin, M. L. *Invest. Ophthalmol.* **8**, 595–604 (1969).
- Bok, D. & Hall, M. O. *J. Cell Biol.* **49**, 664–682 (1971).
- Silverman, M. S. & Hughes, S. E. *Curr. Eye Res.* **9**, 183–191 (1990).
- McNeil, P. L., Muthukrishnan, L., Warder, E. & D'Amore, P. A. *J. Cell Biol.* **109**, 811–822 (1989).
- Finklestein, S. P. et al. *Brain Res.* **460**, 253–259 (1988).
- Baird, A., Mormède, P. & Böhlen, P. *Biochem. biophys. Res. Commun.* **126**, 358–364 (1985).
- Gospodarowicz, D. in *Curr. Top. Dev. Biol.* (ed. Hamilton, M. N.) 57–93 (Academic, San Diego, 1990).
- Adler, R., in *The Retina: A Model for Cell Biology Studies Part I* (eds Adler, R. & Farber, D.) 111–149 (Academic, New York, 1986).
- Hewitt, A. T., Lindsey, J. D., Carbott, D. & Adler, R. *Exp. Eye Res.* **50**, 79–88 (1990).
- Steinberg, R. H. *Exp. Eye Res.* **43**, 695–706 (1986).
- LaVail, M. M., Pinto, L. H. & Yasumura, D. *Invest. Ophthalmol. Vis. Sci.* **21**, 658–668 (1981).
- Schweigerer, L. in *Growth Factors, Differentiation Factors and Cytokines* (ed. Habenicht, A.) 42–55 (Springer, Berlin, 1990).
- Li, L. & Turner, J. E. *Exp. Eye Res.* **47**, 771–785 (1988).

ACKNOWLEDGEMENTS. We thank D. Gospodarowicz for helpful discussions and the gift of aFGF and bFGF, and N. Lawson and G. Riggs for technical and secretarial assistance. This work was supported by the National Eye Institute, the National Retinitis Pigmentosa Foundation, Research to Prevent Blindness, and That Man May See, Inc. M.M.L. is an RPB Senior Scientific Investigator.

A putative serine/threonine protein kinase encoded by the segment-polarity *fused* gene of *Drosophila*

Thomas Pr  at*, Pascal Th  rond,
Claudie Lamour-Isnard, Bernadette Limbourg-Bouchon,
Herv   Tricoire, Inge Erk, Marie-Christine Mariol
& Denise Busson

Centre de G  n  tique Mol  culaire du CNRS, Laboratoire propre associ      l'Universit   P. et M. Curie, F-91198 Gif-sur-Yvette Cedex, France

* Present address: Biology Department, Brandeis University, Waltham, Massachusetts 02254, USA

THE segmented pattern of the *Drosophila* embryo depends on a regulatory cascade involving three main classes of genes¹. An early regulatory programme, set up before cellularization, involves direct transcriptional regulation mediated by gap and pair-rule genes. In a second phase occurring after cellularization, interactions between segment-polarity genes are involved in cell communication^{2,3}. Segment-polarity genes are required for pattern formation in different domains of each metamere and act to define and maintain positional information in each segment⁴. The segment-polarity gene *fused* is maternally required for correct patterning in the posterior part of each embryonic metamere^{1,5}. It is also necessary later in development, as *fused* mutations lead to anomalies of adult cuticular structures and tumorous ovaries. Here we provide molecular evidence that this gene encodes a putative serine/threonine protein kinase, a new function for the product of a segmentation gene. This result provides further insight into segment-polarity interactions and their role in pattern formation.

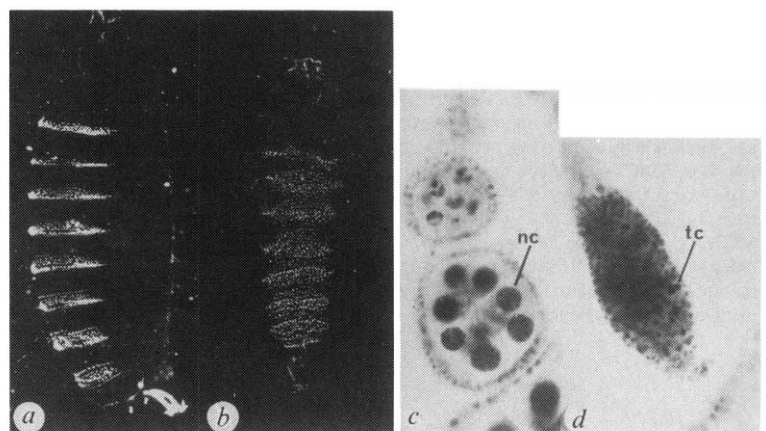
The *fused* (*fu*) gene is one of the 14 segment-polarity genes known to be involved in pattern formation in each segment. In *fu* mutant embryos born from mutant mothers, the posterior part of each segment is deleted and replaced by a mirror-image duplication of the anterior part (Fig. 1a,b). For most alleles, this maternal effect is zygotically rescuable, at least partially, indicating that the gene is active in embryos. In addition, *fu* mutants show adult zygotic effects including wing vein fusion⁶ and ovary anomalies⁷ (Fig. 1c,d). Genetic analysis of 40 *fu* mutants all displaying these different defects, strongly suggested that a single function is involved in both embryonic and adult phenotypes⁸. This view is supported by the fact that a semi-dominant suppressor can suppress the whole *fu* syndrome (T.P., manuscript in preparation).

We have cloned 250 kilobases (kb) of DNA spanning the 17C-D region of the X chromosome where *fu* is located and have assigned this gene to a 15-kb segment defined by the breakpoints of two deletions, *Df(1)fu^{Z4}* and *Df(1)fu^{P1}* (ref. 9) (Fig. 2c). In an initial Southern blot analysis, all *fu* alleles examined were shown to fall into two classes: the 30 viable alleles bore no detectable lesions and the 19 lethal alleles corresponded to very large deletions, the smallest one 40 kb in size. Precise localization was carried out by higher-resolution Southern analysis, to detect lesions as small as five base pairs (bp)¹⁰. Using this method, we found that four viable mutants display small changes of a few bases (Fig. 2a and c). These lesions are distributed through 1,000 bp of DNA belonging to a single transcription unit encoding one 3.2-kb transcript (Fig. 2b,c). This transcript is expressed at a much higher level in females than in males, probably corresponding to its accumulation in oocytes, as shown by its presence in 0-3 h embryos. It is also found in 3-6-h embryos (data not shown). These observations are consistent with the expected expression of *fu*. Finally, P-mediated transformation with a 5.1-kb fragment that only includes this transcription unit rescues the entire embryonic and adult mutant phenotype, proving that it corresponds to the *fu* gene (Fig. 2c).

We have sequenced genomic clones corresponding to the *fu* 3.2-kb transcript (Fig. 3). There is a large open reading frame that encodes a putative polypeptide showing significant similarity with those of the protein kinase family (Fig. 4). The 14 invariant or nearly invariant residues defined by Hanks *et al.*¹¹ from their comparison of 65 different protein kinase catalytic domains, are in the *fu* protein. Protein kinases belong to two different classes depending on the amino acids they phosphorylate: serine/threonine kinases are intracellular, whereas tyrosine kinases are generally transmembrane proteins. Alignment of the putative *fu* catalytic domain with those from other serine/threonine kinases reveals a high degree of amino-acid identity over the entire catalytic domain (Fig. 4): the similarities range from 32% for CDC28 (ref. 12) to 26% for *nim1*⁺ (ref. 13). Sequence comparison has shown that kinases having similar modes of regulation or substrate specificities display similar primary structures. The *fu* catalytic domain cannot easily be included in any of the subfamilies defined by the phylogenetic classification of Hanks *et al.*¹¹ and might define a new serine/threonine kinase subfamily.

Recent reviews have emphasized the role of protein kinases as essential elements of regulatory circuits in differentiated as well as growing cells^{14,15}. Many of these enzymes are involved in transduction of extracellular signals and operate through a cascade of phosphorylation events that amplify and disseminate the effects of a primary signal^{14,16}. The screening of *Drosophila* DNA libraries with vertebrate probes has led to the isolation

FIG. 1 Embryonic and adult aspects of the *fu* phenotype. a, b, Dark-field micrographs showing ventral cuticular pattern of wild-type (a) and *fu^A* embryos (b). In *fu^A* embryos born from *fu^A* mothers, the whole naked cuticle of the posterior part of each abdominal and thoracic segment is absent and replaced by a mirror-image duplication of the anterior part. Denticles of the duplicated belts show reverse polarity. The Keilin's organs, which mark the anteroposterior border in thoracic segments are absent in *fu* embryos but the segmental boundaries are clearly visible. Head structures corresponding to the cephalo-pharyngeal skeleton are abnormally positioned. Terminal structures such as anal pads, tuft, posterior spiracles and Filzk  rper are normal. c, d, Preparation of egg-chambers from wild-type (c) and *fu¹* females (d). Wild-type egg-chambers presented here correspond to stages 4-6 of King³⁴. Each contains 15 polyploid nurse cells (nc), which provide the oocyte with maternal components. Up to 50% of the egg-chambers from *fu¹* females are tumorous. One chamber can contain as many as 10,000 tumorous cells (tc).



METHODS. The origins and genetic properties of the deletions and viable *fu* alleles are described in ref. 8. The fragment used for P-mediated germ-line transformation was cloned into the pW6 vector and injected into *w¹¹¹⁸*; *P(ry⁺A2-3)* *Sb/TM6 Ubx* mutant embryos³⁵. Insertions from independent transformed lines were transferred by appropriate crosses into *fu¹* and *fu^A* mutant strains that were then examined for their phenotypic rescue. The technique used for Southern analysis performed to detect small changes (<100 bp) is described in ref. 10. Poly(A)⁺ RNAs were fractionated on formaldehyde agarose gels, transferred to nylon filters and hybridized with ³²P-labelled probes generated by random-priming. The orientation of transcription was deduced using single-stranded RNA probes and by DNA sequencing of genomic and complementary DNA clones.

METHODS. The nucleotide sequence was determined for both strands by dideoxy-sequencing³⁶ using sequenase (USB). Genomic DNA restriction fragments were cloned into M13 mp18 and M13 mp19 vectors and sequenced using synthetic oligonucleotide primers.

a

b

c

d

AGATCTCGCGCAGATGGCGCACTGATAGGAAAGAAGAGCCGCAACCTGGGATGCGGATTGGGAGTGTGGTCAATGGCAGTGGGAATCCG
 AACTCAATTAGACTCGCGCTGGGAAGCGGAAGCGGTGGCAGAAATGTCTCGCGGTGGCGGAACTCACCAGTCAGCGCGTGGGACA
 GATCAGTACAGGTTGGGCTCGCGCACTCGTGTGGGATTGGGTTGTGTCTCTCACTGGTGGCTGTGTCTCAATTGGCAATCATTTAT
 TTTCACAAAGATTATCTCGCTGGCGCAAAATAAAAACGAGCCAGCAATCAGCTGTCTCTCGCGCGTGTGTCTAGGSGTACCGTAC
 AGTTGCTGTCAAAAGGTCATTGTGTGTGGTTGGTCCACATGAATTTAGGTACATACAGATATAGTACATACATATATAGTAT
 TATTATCTACTGCATCTTTGGAAATTTTAAATTTATATATATGCAATCATTTATTGCAATTTATTAAGTAAAAATGTACATTAAGA
 ATATACATATAAATTTTATATATATATATATATATATAGTTAAAGTTCGTTCCAACTTTTATCACTTAAATATATATATATGCA
 ATACATTAATACCTACTGCCACAGATAAAAAGCGCGAATCTAGTCTGTATTGTCTGTGGAATGTATCATATCTAGTATTTTCA
 GTACAAATGCTTTCCAGAAAAGCGTCAATTATACAGAAAAGAAACCGCTGTATGAGCAATCTAGTATTTTCAATACCTACCTACCGA
 CGATTTCCCACTCGGAGTCAACCTCTACGCTCATACCGCAGCACTTCAGTGGTAATCAGTCCCAATCCGACGAAAACAAAGAA
 ACCATGAAACCGCTACGCGGTAACTCGCTCGTGGGCGCAAGCATCTCTCGGTTGGCTATACAGAGGAGCAACGCAAGGACACAGCAAGT
 MetAsnArgTyrAlaValSerSerLeuValGlyGlnGlySerPheGlyCysValTyrLysAlaThrArgLysAspAspSerLysVal
 ValAlaIleLysValIleSerLys
 ctgcagCGCGAAGAGCCCAAGAGAGCTGAAGAAATTGGCGAGGAGTGGCAGATTCAGGCGCGGCTGAAGCATCCGCGCTCATCG
 ArgGlyArgAlaThrLysGluLeuLysAsnMetLysArgGlyCysAspIleGlnAlaArgLysLysHisProHisValIleGln
 ATGATCGAGTCTTCGAGTCAAGAGCAAGCACTTTCTGGTCTCACTGAGTTCGCGCTGATGGAATCGCAGCGCTACTGCTCTCAATG
 MetIleGluSerPheGluSerLysThrAspLeuPheValThrGlnPheAlaLeuMetAspLeuHisArgTyrLysSerTyrAsnGln
 GCCATGGCGCAGGAGCGCGCACTCGGCTGACCGCGCATCTGCTGTCTGCTACTACCTGCATTCAAACCCATCTCCACCGGAG
 AlaMetGlyGluGluProAlaArgArgValThrGlyHisLeuValSerAlaLeuTyrTyrLysHisSerAsnArgIleLeuHisArgAla
 CTCAAACCGCAAACTCTGCTCGCAGCAAGCAATGCAGCAAACTCTGCGCAATTTGGATGCGCGCGCAACATGACCTGGTACCC
 LeuLysProGlnAsnValLeuLeuAspLysAsnMetHisAlaLysLeuCysAspPheGlyLeuAlaArgAsnMetThrLeuGlyThrHis
 GTGCTCACTCTGATGAGGGAACCGCCCTCTACATGGCGCGGAGTCTGGCGAGAGGCCATACGACCATCTGCGCAGATGTGTG
 ValLeuThrSerLysLysGlyThrProLeuTyrMetAlaProGluLeuLeuAlaAspGluProTyrAspHisHisAlaAspMetThrTrp
 CTGGGCTGATACGCTCAAGAAAGCAAGCGCGTCAAGCGCCCTCTGCTGCAAGTCCATCTGCTATCTGTGAAGATGATCAAGCAAG
 LeuGlyCysIleAlaTyrGluSerMetAlaGlyGlnProProPheCysAlaSerSerIleLeuHisLeuValLysMetIleLysHisGln
 GACGTGAAGTGGCGAGCAGCTGACTTGCAGATGCCGCTCTCTCTCTCAAGGCGCTCTGTGAGGAAGGACCGCGCTGCGCATATCT
 AspValLysTrpProSerThrLeuLeuThrCysGlnCysArgSerPheLeuGlnGlyLeuLeuLysAspProGlyLeuArgIleSerThr
 ACGCAGCTGTGTGTGACCCCTTCGTTGAGGACGCAATCTTTATGCGAGAAACGAGCGGAGCGGCGGCAAGGAATCGCCTTTC


```

fused  /.....domain 1...../.....domain 2...../.....domain 3.
cAMPK  HNRIVASSLVGGGFCGYKA---TRKDDSKVVAIVISK-RGRATKELNLRRCEDIOA
PKC    LDPERKTLTGKSGFQNVIV---CHRPKIDYVAKKLLDQKRVVVKQVHTLNKRKILQ
CaMK   LDDNFNMLKSGSGFQNVIA---DRKGTELVAKKLLKDDVYVQDOVCTWIKRVL
nim1   TDYVLYEDIKLGSFVVRRC---VKLCGTGVEYAKKILNT-KKLSARDHOKTEERARIC
CDC28  JGVHRLGKTLTGKTSCTVIA---KHAKTGDLAAIKIPI-R---YASIGHEILMMR
      LANYKRLKGVVGGTGVYVAKALDRPGQGVVYVAKIRL-ESEDEGVSTAIRISLLK
      * * * * *

fused  /.....domain 5...../.....domain 6...../.....domain 7...../.....domain 8
cAMPK  RL-KHPHYVTEIESFESKTD-LEFVVE-AL-NDLHY---LSYNGANGEEPARRVTHGL
PKC    AI-OFPEFLYSRLVHFEDNSM-LVMVLEKVG-CGPFSS---LKKVGRFSEPHSRFVAQI
CaMK   LLDKPPFLTQLHSCFQTVDR-LYFVPEY-VN-CGDLMYH---IQGVKFKPEQAVFVAARI
nim1   LL-KHSNIVRLHDSISEEGF-HYFVFDL-VTGGELPED---IVAREVYSEADASHCQOI
CDC28  LL-RHPNIRLRLVDVWTDHQR-MYLALEYVDP-GELFHY---IRKHGPLSEADASHCQOI
      EL-KDDNIVRLVDIVSHDAHKLYLVFEE-LD-LDLKRYMEGIPKDOPLADIVKFFMMOL
      * * * * *

fused  /.....domain 9...../.....domain 10.....
cAMPK  VSALYYLHNRILHRDLKPPVLL---DK-NHARLDTGLARNHTLGLTVLSIKGTP
PKC    VI-AFEYLYLHDLVLDLKPENLLI---DS-OCYKLVDFGAKRVKGRVTLGCS---TPE
CaMK   SIGLFFLLKRGIVLDLKPENLLI---DS-EGHKTIDFGCKEHNMDGVTTTRTCGTEG
nim1   LEAVLHCHMGVYVDRDLKPENLLI---DS-EGHKTIDFGCKEHNMDGVTTTRTCGTEG
CDC28  LDVAHCHMGRFPHRDLKPPVLLI---KVNQQKIDFGCA-TVEPNDSCLENYCGSLH
      CKGIAYCHSHRILHRDLKPENLLI---NK-DGNLKLDFGLARAFGVPILRAYTHVLTSLW
      * * * * *

fused  /.....domain 11...../.....domain 12...../.....domain 13.....
cAMPK  YLAPE-LLADEPV-DHAAHMSLFCIAYSMAGOPP-ECA-SSILHLVKMIK---HEDV
PKC    YLAPE-LILSKGY-NKAVDMALVLYVMAAGYVE-FFA-DOPIQIVEKIV---SGKV
CaMK   YLAPE-LIAYOPV-CKSVDMALVLYVMAAGYVE-FFA-DOPIQIVEKIV---EHNV
nim1   YLAPE-VLRKEAY-CKPVDMALVLYVMAAGYVE-FFA-DOPIQIVEKIV---AGAY
CDC28  YLAPE-LVSHKPY-GAPADMSGVLYVLYVMAAGYVE-FFA-DOPIQIVEKIV---HGA
      YLAPE-LVGGKPY-STGVDMALVLYVMAAGYVE-FFA-DOPIQIVEKIV---HGA
      * * * * *

fused  /.....domain 14...../.....domain 15...../.....domain 16.....
cAMPK  KWFS---TLTCCBSFLGKLEKCPGIRISMTOLLCHFFVEGRIFIAETQAEAA
PKC    RFTS---HFGSDLDKLLRNLDVLTTRNGNLKAGVNDIKNOKFASTOWIAFF
CaMK   SYPK---SLSKVAYSLCKGLTTHGKGRGCGPEGRDVRHAFERRIDWEKLE
nim1   DHPSPWDVTPPAKNLQMLTINPAKRTIAHEALKHFFVYVCCSTVASMHRQE
CDC28  DIPS---SISSAODLLHRLMDVNSPTITPFFFSHPFLMGCTSLSSMDSTTP
      IMEDIVYLPDPRGIDLLKLLAYDPIHRTISARRAAIHPEYFQESGTGVVYKALD
      * * * * *

```

FIG. 4 Sequence comparison of *fu* and other serine/threonine kinase catalytic domains (single-letter amino-acid code). cAMPK, Cyclic AMP-dependent protein kinase catalytic subunit from *D. melanogaster*¹⁷; PKC, protein kinase C α -form from bovine brain³⁷; CaMK, calcium-calmodulin-dependent protein kinase type II, β -subunit from rat brain³⁸; *nim1*, 'new inducer of mitosis' gene product from *Schizosaccharomyces pombe*¹³; CDC28, 'cell division cycle' gene product from *Saccharomyces cerevisiae*¹². These five proteins are the representatives of five different subfamilies in the serine/threonine kinase family. Division into sub-domains I-XI is according to Hanks *et al.*¹¹. Gaps, represented by dashes, have been introduced to optimize the alignment. ●, Omission of 21 amino acids in the CDC28 sequence. *, Fourteen invariant or nearly invariant residues characteristic of all kinases: the lysine residue in subdomain II is involved in phosphotransfer reactions, the invariant residues in subdomains I, VI and VII are implicated in ATP-binding; role of invariant residues in subdomains III, V, IX and XI is unknown. The two consensus sequences, DLKPEN in subdomain VI and GT/SXXY/FXAPE in subdomain VIII, indicators of serine/threonine specificity, are underlined by crosses. Residues identical in all six sequences are in bold type and boxed. Conservative residues when compared with those of the *fu* sequence are in bold type and underlined. Conservative amino-acid substitutions were determined according to the following groupings: ILVM, FYW, DEQN, KRH, GAPST and C. The statistical significance of the comparison of the *fu* sequence with any other sequence was estimated by the RDF program of Lipmann and Pearson³⁹. The Z values were 37(cAMPK), 46(PKC), 43(CaMK), 49(nim1) and 45(CDC28). A Z-value >3 is believed to be biologically significant³⁹.

of many sequences encoding putative protein kinases^{11,14,17-21}, but there are only a few cases where a specific mutant phenotype has been associated with the loss of function of a particular kinase^{20,22-24}. Our observations strongly suggest that the *fu* mutant phenotype arises from lack of a serine/threonine kinase activity.

Many serine/threonine kinases are implicated in cell-cycle control in yeasts and in mammals¹⁴⁻¹⁶. A *Drosophila* gene homologous to the human *c-raf1* oncogene affects the normal rate of proliferation of both somatic and germinal cells²⁰. The observation that *fu* mutant females display tumorous egg-chambers with several hundreds of proliferating cells (Fig. 1d) supports the hypothesis that the *fu* protein might be involved in control of cell division, at least during ovarian development.

Embryos lacking the *fu* product die, presenting a segment-polarity phenotype. So far, *fu* is the only segmentation gene that could display serine/threonine kinase activity. Several products of developmental genes, including those of the segmentation genes *Krüppel* (*Kr*)²⁵, *fushitarazu* (*ftz*)²⁶ and *engrailed* (*en*)²⁷, are phosphorylated by serine/threonine kinases. For *ftz*, different forms of the modified protein can be found at different stages of embryonic development, suggesting a spatiotemporal modulation of its activity by phosphorylation²⁶. It is likely that post-translational modifications play an important part during the segmentation process, especially after cellularization, when segment-polarity gene products are thought to be involved in cell-cell communications⁴. Segment-polarity genes encode a large array of products, including DNA-binding proteins (*en*²⁸, *gooseberry*²⁹) and proteins that can be directly involved in signal transduction (*wingless* product is homologous to a secretory protein^{30,31} and *patched* is a transmembrane protein^{32,40}). One putative substrate for the kinase encoded by *fu* is the *en* product; *fused* activity is required in the posterior part of each embryonic segment where the *en* protein is located. In addition, the viable *en* phenotype is strongly enhanced in flies mutant for both *fu* and *en*³³, suggesting that the *fu* product interferes with the expression of the *en* function. The kinase encoded by *fu* might thus be an important intermediate in the signal transduction chain through which segment-polarity products control pattern formation in each embryonic segment.

Note added in proof: The segment polarity gene *zeste-white3* (*shaggy*) described recently, encodes proteins that also have homology to serine/threonine protein kinases^{41,42}. □

Received 29 May; accepted 21 June 1990.

- Nusslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
- Ingham, P. *Nature* **335**, 25-34 (1988).
- Di Nardo, S., Sher, E., Heemskerk-Jongens, J., Kassis, J. A. & O'Farrell, P. H. *Nature* **332**, 604-609 (1988).
- Martinez-Arias, A., Baker, N. E. & Ingham, P. W. *Development* **103**, 157-170 (1988).
- Martinez-Arias, A. *J. Embryol. exp. Morphol.* **87**, 99-114 (1985).
- Fausto-Sterling, A. *Dev. Biol.* **63**, 358-369 (1978).
- Smith, P. A. & King, R. C. *J. natn. Cancer Inst.* **36**, 445-461 (1966).
- Bussion, D., Limbourg-Bouchon, B., Mariol, M. C., Pr  at, T. & Lamour-Isnard, C. *Wilhelm Roux Arch. dev. Biol.* **197**, 221-230 (1988).
- Mariol, M. C., Pr  at, T. & Limbourg-Bouchon, B. *Molec. cell. Biol.* **7**, 3244-3251 (1987).
- Pr  at, T. *Nucleic Acids Res.* **18**, 1073 (1990).
- Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42-52 (1988).
- Lorincz, A. T. & Reed, S. I. *Nature* **307**, 183-185 (1984).
- Russel, P. & Nurse, P. *Cell* **49**, 569-576 (1987).
- Hunter, T. *Cell* **50**, 823-829 (1987).
- Dunphy, W. G. & Newport, J. W. *Cell* **55**, 925-928 (1988).
- Lee, M. & Nurse, P. *Trends Genet.* **4**, 287-290 (1988).
- Kalderon, D. & Rubin, G. M. *Genes Dev.* **2**, 1539-1556 (1988).
- Schaeffer, E. *et al. Cell* **57**, 403-412 (1989).
- Mark, G. E. *et al. Molec. cell. Biol.* **7**, 2134-2140 (1987).
- Nishida, Y. *et al. EMBO J.* **7**, 775-781 (1988).
- Shilo, B. Z. *Trends Genet.* **3**, 69-72 (1987).
- Montell, C. & Rubin, G. *Cell* **52**, 757-772 (1987).
- Hafen, E., Basler, K., Edstroem, J. E. & Rubin, G. M. *Science* **236**, 55-63 (1987).
- Sprenger, F., Stevens, L. M. & Nusslein-Volhard, C. *Nature* **338**, 478-483 (1989).
- Oillo, R. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5700-5704 (1987).
- Krause, H. M. & Gehring, W. J. *EMBO J.* **8**, 1197-1204 (1989).
- Gay, N. J., Poole, S. J. & Kornberg, T. B. *Nucleic Acids Res.* **16**, 6637-6647 (1988).
- Jose, A., McGinnis, W. J. & Gehring, W. J. *Nature* **313**, 284-289 (1985).
- Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. *Cell* **47**, 1033-1040 (1986).
- Rijsewijk, F. *et al. Cell* **50**, 649-657 (1987).
- Van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. A. *Cell* **59**, 739-749 (1989).
- Nakano, Y. *et al. Nature* **341**, 508-513 (1989).
- Fausto-Sterling, A. & Smith-Schless, H. *Genetics* **101**, 71-80 (1982).
- King, R. C. *Ovarian Development in Drosophila melanogaster* (Academic, New York, 1970).
- Robertson, H. M. *et al. Genetics* **118**, 461-470 (1988).
- Sanger, F., Nicklen, S. & Coulson, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Parker, P. J. *et al. Science* **233**, 853-859 (1986).
- Bennet, M. K. & Kennedy, M. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1794-1798 (1987).
- Lipman, D. J. & Pearson, W. R. *Science* **227**, 1435-1441 (1985).
- Hooper, J. E. & Scott, M. P. *Cell* **59**, 751-765 (1989).
- Siegfried, E., Perkins, L. A., Capaci, T. M. & Perrimon, N. *Nature* **345**, 825-829 (1990).
- Bontouis, M. *et al. EMBO J.* (in the press).

ACKNOWLEDGEMENTS. We thank our colleagues for discussions and critical reading of the manuscript. Transformation experiments were performed with the help of J. Collignon and E. Fabre, students of the Molecular and Cellular Genetic DEA of Paris VI and Paris XI Universities and V. Vincent, Paris XI University student. This work was supported partly by la Ligue Nationale Fran  aise contre le Cancer and la Fondation pour la Recherche Medicale. T.P. was the recipient of an ARC fellowship.

Megabase inversions in the human genome as physiological events

Gottfried M. Weichhold, H.-Gustav Klobeck, Rita Ohnheiser, Gabriele Combriato & Hans G. Zachau*

Institut für Physiologische Chemie der Universität München, Goethestrasse 33, D-8000 München 2, FRG
* To whom correspondence should be addressed

THE genes of the immunoglobulin κ light chains are assembled during B-cell differentiation by somatic recombination of one of the $V\kappa$ (variable) gene segments and the $J\kappa$ - $C\kappa$ (joining-constant) gene region. This seems to occur by deletion of the DNA between $V\kappa$ and $J\kappa$ - $C\kappa$ if they are arranged in germ-line DNA in the same transcriptional polarity or by inversion of a fragment containing the $V\kappa$ gene if the polarities are opposite (Fig. 1a; see refs 1-3 for reviews; refs 4-8). We have cloned 75 $V\kappa$ genes and pseudogenes of the human κ locus and linked them in large contigs (refs 3, 9 for reviews). There seem to be no more than 85 such genes, less than 50 of these being potentially functional¹⁰. Thirty-eight of the cloned genes have the same transcriptional polarity as $J\kappa$ - $C\kappa$ and are part of the so-called $J\kappa$ proximal cluster; 35 genes in a distal cluster (the result of a duplication event in evolution) have a polarity that was suggested to be opposite to the one of $J\kappa$ - $C\kappa$ (refs 9, 11). We now show that the $V\kappa$ genes of the proximal cluster rearrange by a deletion mechanism whereas the others join $J\kappa$ - $C\kappa$ by inversion of megabase-sized DNA fragments.

For this study we selected two B cell lines that produce κ chains, RPMI6410 and GM607, which both contain one germ-line allele (κ^0) and one $V\kappa$ gene rearranged to $J\kappa 1$ (κ^+) (refs 12-14). In RPMI6410 it is the gene A17 of the $J\kappa$ proximal cluster of the locus and in GM607 the A3 gene of the distal cluster. The evidence for the proposed rearrangement mechanisms rests on sequence analysis of the signal joint in GM607 that contains the flanking sequences of the A3 gene and of $J\kappa 1$ (Fig. 1b) and on mapping experiments with pulsed field gel electrophoresis (PFGE; Fig. 2). The germ-line situation of the

TABLE 1 κ locus-derived fragments of the three cell-line DNAs

	RPMI6410		PC-3		GM607
NotI*	0.87 1-4	$p\kappa^0$	1.05 1-4	$p\kappa^0$	1.15 1-4 $p\kappa^0$
	0.64 3-5	$d\kappa^0$	0.85 3-5	$d\kappa^0$	1.03 3-5 $d\kappa^0$
	0.83 1,4	$p\kappa^+$	1.28 1-4	$p\kappa^0$	0.83 1,4 $p\kappa^+$
	0.98 3-5	$d\kappa^+$			1.25 2-5 $d\kappa^+$
NotI-NruI*	0.87 1-4	$p\kappa^0$	1.05 1-4	$p\kappa^0$	1.05 1-4 $p\kappa^0$
	0.64 3-5	$d\kappa^0$	0.85 3-5	$d\kappa^0$	0.85 3-5 $d\kappa^0$
	0.64 1,4	$p\kappa^+$			0.64 1,4 $p\kappa^+$
	0.85 3-5	$d\kappa^+$			1.17 2-5 $d\kappa^+$
NruI†	1.5 1-4	$p\kappa^0$	1.5 1-4	$p\kappa^0$	1.5 1-4 $p\kappa^0$
	1.6 3-5	$d\kappa^0$	1.6 3-5	$d\kappa^0$	1.30 3-5 $d\kappa^0$
	1.30 3-5	$d\kappa^+$			1.20 1-4 $p\kappa^+$
	1.20 1,4	$p\kappa^+$			1.6 2-5 $d\kappa^+$
MluI‡	1.04 1-4	$p\kappa^0$	1.04 1-4	$p\kappa^0$	1.04 1-4 $p\kappa^0$
	0.42 3-5	$d\kappa^0$	0.42 3-5	$d\kappa^0$	0.42 3-5 $d\kappa^0$
	0.62 3-5	$d\kappa^+$			0.64 1,4 $p\kappa^+$
	0.64 1,4	$p\kappa^+$			0.82 2-5 $d\kappa^+$

For each fragment the size (Mb), the hybridization signals with probes 1-5 (Fig. 3) and the resulting assignment to the proximal (p) or distal (d) copy of the κ^0 or κ^+ allele is given.

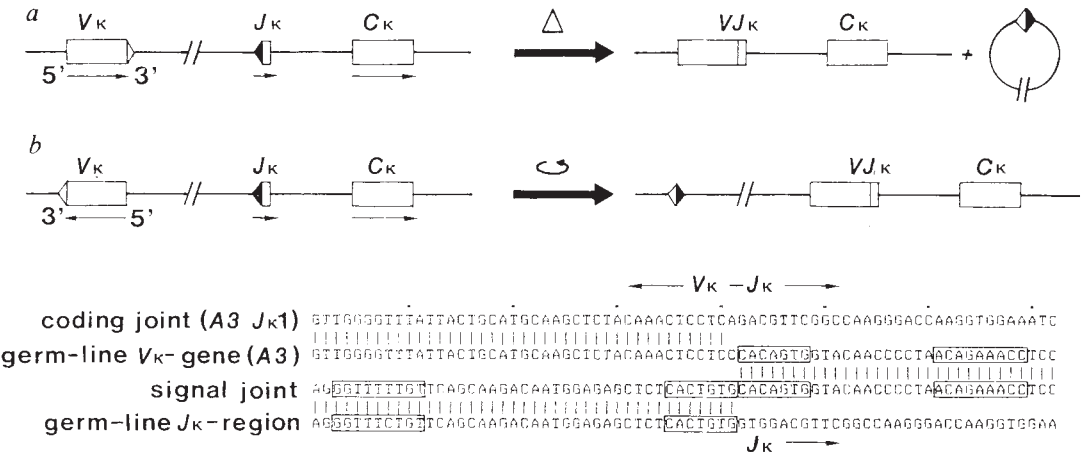
* The sizes of the *NotI* and *NotI-NruI* fragments are best considered by focusing on the four *NotI* sites marked by asterisks in Fig. 3b. The two *NotI* sites that are inseparable from *NruI* sites are accessible only in PC-3, fully in the distal copy and partially in the proximal copy. The distances between the *NotI-NruI* sites and the central *NotI* sites differ between the cell lines by 50-100 kilobases (kb). The other two *NotI* sites marked by asterisks are accessible only in the κ^0 allele of RPMI6410. The *NotI* and *NotI-NruI* fragments in $d\kappa^+$ of GM607 are 0.1 Mb shorter than the $d\kappa$ fragments of the other cell lines, which may be due to a 0.1-Mb deletion or to an additional *NotI* site.

† The size difference of the *NruI* fragments derived from the $d\kappa^0$ alleles (1.6 versus 1.3 Mb) which is also seen in DNAs from several individuals is caused by the presence (or accessibility) of an additional *NruI* site at the 3' side. The proximal alleles vary in size by 0.1 Mb between cell lines and individuals which explains the 1.2-Mb instead of a 1.1-Mb *NruI* fragment for $p\kappa^+$ of RPMI6410.

‡ One of the *MluI* sites marked by an asterisk in Fig. 3 is absent/inaccessible in either the $d\kappa^0$ or the $d\kappa^+$ allele of RPMI6410, giving rise to a 0.62-Mb fragment in addition to the 0.42-Mb fragment.

κ locus was mapped in the cell line PC-3. The structure of the κ locus in the two alleles of PC-3 and in the κ^0 alleles of GM607 and RPMI6410 is largely identical to the one in the contigs that had been cloned from the DNAs of various individuals^{3,9}. The

FIG. 1 Coding and signal joints in immunoglobulin κ gene rearrangements. a, Open boxes, gene segments; triangles, the hepta- and nonanucleotide signal sequences. In the deletion reaction (Δ) the signal joint is excised while in inversions ($\odot$) it is retained in the genome. b, The sequence of the signal joint, which was determined in subclones of the cosmid clone cos607/2 (ref. 16) is deposited in the EMBL data library, Heidelberg (accession number X54247), together with 1,094 bp of 5' and 73 bp of



gene A5 is identical to the one of cos142 of the so-called Aa region¹³ of the distal $V\kappa$ gene cluster. The sequence of the signal joint is compared with the ones of the A3 and $J\kappa 1$ gene flanks^{13,20}. The coding joint, that is, the rearranged $V\kappa$ gene V607 of the cell line GM607 (ref. 14) is included for comparison with the progenitor germ-line gene A3.

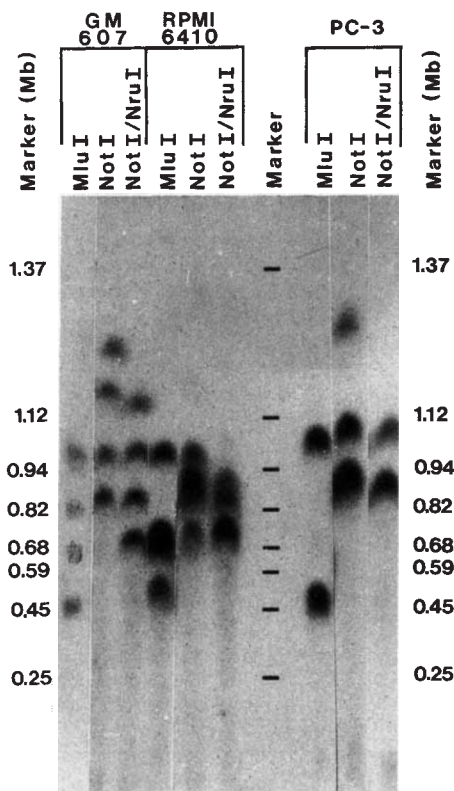


FIG. 2 DNA fragments derived from the κ loci of the cell lines GM607 (κ^+ with inversion, κ^0), RPMI6410 (κ^+ with deletion, κ^0) and PC-3 (κ^0 , κ^0). The PFGE was run in a Rotaphor²¹ with V2 rotor in 0.8% agarose at 14 °C with 120 V for 45 h with 330 s/cycles and then 17 h with 220 s/cycles. *Saccharomyces cerevisiae* (X2180-1B) chromosomes were used as markers, shown in Mb (ref. 22, and R. Anand, personal communication). The blot was hybridized with the probe m143-1 (probe 4 in Fig. 3).

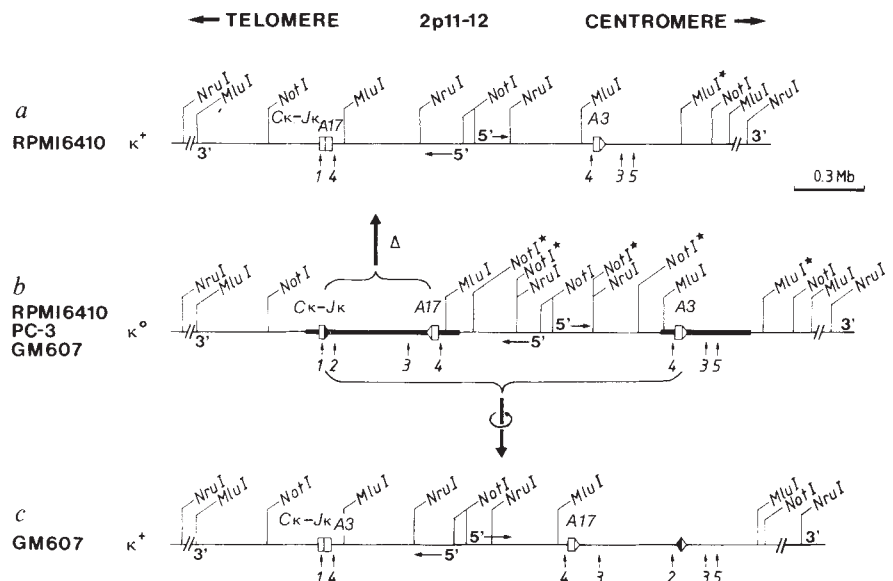
locus has been mapped in PC-3 with 11 rare-cutting nucleases and in RPMI6410 and GM607 with 7 of them (ref. 11; G.M.W., unpublished observations). Eleven hybridization probes were used. The argument concerning deletion and inversion can be made, however, using only five probes and digests with three nucleases and one nuclease combination.

Figure 2 shows a representative blot hybridization experiment. Table 1 presents the sizes, hybridization properties and assignments of the fragments that the restriction maps (Fig. 3) are based on. Three hybridization probes are unique (1, 2 and 5 in Fig. 3), the others hybridize to both duplicated regions (for example, 3 and 4 in Fig. 3). The PFGE fragments were assigned to the $J\kappa$ proximal or to the distal $V\kappa$ gene cluster (p and d in Table 1) on the basis of hybridizing the blots consecutively to five to nine different probes. The κ locus and its surrounding

mapped areas comprise about 3.5 megabases (Mb) and the locus itself slightly less than 2 Mb.

The deletion of the DNA between $A17$ and $J\kappa 1$ is most clearly seen in the $MluI$ digests: the $J\kappa$ proximal fragment of RPMI6410 DNA loses the expected 0.4 Mb and no longer hybridizes to probes 2 and 3 (Table 1; Fig. 3a, b). Also, the inversion is best demonstrated in the $MluI$ digests: the proximal fragment of GM607 DNA decreases and the distal one increases in length by 0.4 Mb and the $J\kappa 1$ flank (probe 2) becomes part of the signal joint in the distal fragment (Table 1; Fig. 3b, c). In GM607, the inverted fragment has a size of about 1.6 Mb. As the $V\kappa$ genes of the distal $V\kappa$ contig have the same 5'-3' polarity^{11,13} all of them must rearrange by inversion of fragments that are somewhat smaller or larger depending on the relative location of the respective gene in the cluster. This is also true for the $A2$

FIG. 3 Restriction maps of the κ^+ alleles of the cell lines RPMI6410 (with deletion) and GM607 (with inversion) and a composite map of the κ^0 alleles of RPMI6410, GM607 and PC-3. The maps, which were constructed on the basis of the data of Table 1, are consistent with numerous additional cleavages and hybridizations (see text). Asterisks: $NotI$ and $MluI$ sites that are not or only partially cleaved in certain alleles (footnotes to Table 1). The cloned regions are drawn as thick lines in the PC-3 map; interruptions in the $C\kappa$ -containing contig^{3,9} are not shown. Gene segments and the signal joint are drawn as in Fig. 1a (not to scale). The slashes on both sides of the maps represent stretches of DNA without known restriction sites. Sites of hybridization are shown for the following probes: probe 1, mC-2 is the 2.6-kb *EcoRI* fragment containing $C\kappa$ (ref. 14); 2, AF-2/12 (ref. 23); 3, m142-2 (ref. 13); 4, m143-1 (ref. 24); 5, m656-1 (ref. 13).



gene (located adjacent to A3) that was recently shown to code for the V_k part of a functional antibody¹⁵.

Inversion of relatively small DNA fragments occur in the human κ locus^{6,16,17} and, for example, in the T-cell-receptor β -chain locus¹⁸. Inversions also occur in the mouse κ locus^{5,19} but as this locus has not been mapped in detail the sizes of the inverting fragments are not known. Therefore, the present example is the first demonstration of a physiologically occurring megabase-sized inversion leading to a functional product.

Do megabase inversions confer a functional advantage to the cell? They probably do, at least in comparison with megabase

deletions that would be required if the distal V_k genes had the same polarity as J_k - C_k . Many V_k genes and possibly also genes with other functions that may be located in the 0.8 Mb between the two V_k clusters (Fig. 3b) would be lost in such deletions. As V_k - J_k rearrangements frequently lead to aberrant coding joints that can be replaced in subsequent rearrangements of another V_k to a still unused J_k gene, retaining the V_k genes in inversions should be advantageous to the cell. This does not provide an answer to the question of why the κ locus is so large, but it makes it seem reasonable that megabase inversions are a good way for the cell to handle the mechanics of a large locus. □

Received 13 June; accepted 6 July 1990.

1. Baltimore, D. *Nature* **319**, 12-13 (1986).
2. Lewis, S. & Gellert, M. *Cell* **59**, 585-588 (1989).
3. Zachau, H. G. in *The Immunoglobulin Genes* (eds Honjo, T., Alt, F. W. & Rabbitts, T.) 91-109 (Academic, London, 1989).
4. Steinmetz, M., Altenburger, W. & Zachau, H. G. *Nucleic Acids Res.* **8**, 1709-1720 (1980).
5. Van Ness, B. G., Coleclough, C., Perry, R. P. & Weigert, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 262-266 (1982).
6. Deev, S. M., Combrato, G., Klobbeck, H.-G. & Zachau, H. G. *Nucleic Acids Res.* **15**, 1-14 (1987).
7. Lewis, S., Gifford, A. & Baltimore, D. *Nature* **308**, 425-428 (1984).
8. Hesse, J. E., Lieber, M. R., Gellert, M. & Mizuuchi, K. *Cell* **49**, 775-783 (1987).
9. Zachau, H. G. *Biol. Chem. Hoppe-Seyler* **371**, 1-6 (1990).
10. Meindl, A., Klobbeck, H.-G., Ohnheiser, R. & Zachau, H. G. *Eur. J. Immun.* **20**, 1855-1863 (1990).
11. Lorenz, W., Straubinger, B. & Zachau, H. G. *Nucleic Acids Res.* **15**, 9667-9676 (1987).
12. Klobbeck, H.-G., Meindl, A., Combrato, G., Solomon, A. & Zachau, H. G. *Nucleic Acids Res.* **13**, 6499-6513 (1985).
13. Straubinger, B. et al. *J. molec. Biol.* **199**, 23-34 (1988).

14. Klobbeck, H.-G., Solomon, A. & Zachau, H. G. *Nature* **309**, 73-76 (1984).
15. Scott, M. G. et al. *J. Immun.* **143**, 4110-4116 (1989).
16. Klobbeck, H.-G., Zimmer, F.-J., Combrato, G. & Zachau, H. G. *Nucleic Acids Res.* **15**, 9655-9665 (1987).
17. Lorenz, W., Schable, K. F., Thiebe, R., Stavnezer, J. & Zachau, H. G. *Molec. Immun.* **25**, 479-484 (1988).
18. Malissen, M. et al. *Nature* **319**, 28-33 (1986).
19. Feddersen, R. & Van Ness, B. *Nucleic Acids Res.* **17**, 9797-9809 (1989).
20. Hieter, P. A., Maizel, J. V. Jr & Leder, P. *J. biol. Chem.* **257**, 1516-1522 (1982).
21. Ziegler, A., Geiger, K.-H., Ragousis, J. & Szalay, G. J. *Clin. Chem. clin. Biochem.* **25**, 578-579 (1987).
22. Anand, R. *Trends Genet.* **2**, 278-283 (1986).
23. Klobbeck, H.-G., Combrato, G. & Zachau, H. G. *Biol. Chem. Hoppe-Seyler* **370**, 1007-1012 (1989).
24. Pohlenz, H. D. et al. *J. molec. Biol.* **193**, 241-253 (1987).

ACKNOWLEDGEMENTS. The work was supported by Bundesministerium für Forschung und Technologie (Genzentrum München) and Fonds der Chemischen Industrie.

Inhibition of HIV replication by pokeweed antiviral protein targeted to CD4⁺ cells by monoclonal antibodies

Joyce M. Zarling*, Patricia A. Moran*, Omar Haffar*, Joan Sias*, Douglas D. Richman†, Celsa A. Spina†, Dorothea E. Myers‡, Virginia Kuebelbeck‡, Jeffrey A. Ledbetter* & Fatih M. Uckun‡

* Oncogen, 3005 1st Avenue, Seattle, Washington 98121, USA

† University of California, San Diego and Veterans Administration Medical Center, San Diego, California 92161, USA

‡ Tumor Immunology Laboratory, Department of Therapeutic Radiology, University of Minnesota, Minneapolis, Minnesota 55455, USA

FUNCTIONAL impairment and selective depletion of CD4⁺ T cells, the hallmark of AIDS, are at least partly caused by human immunodeficiency virus (HIV-1) type 1 binding to the CD4 molecule and infecting CD4⁺ cells^{1,2}. It may, therefore, be of therapeutic value to target an antiviral agent to CD4⁺ cells to prevent infection and to inhibit HIV-1 production in patients' CD4⁺ cells which contain proviral DNA^{3,4}. We report here that HIV-1 replication in normal primary CD4⁺ T cells can be inhibited by pokeweed antiviral protein, a plant protein of relative molecular mass 30,000 (ref. 5), which inhibits replication of certain plant RNA viruses⁶⁻⁸, and of herpes simplex virus, poliovirus and influenza virus⁹⁻¹¹. Targeting pokeweed antiviral protein to CD4⁺ T cells by conjugating it to monoclonal antibodies reactive with CD5, CD7 or CD4 expressed on CD4⁺ cells, increased its anti-HIV potency up to 1,000-fold. HIV-1 replication is inhibited at picomolar concentrations of conjugates of pokeweed antiviral protein and monoclonal antibodies, which do not inhibit proliferation of normal CD4⁺ T cells or CD4-dependent responses. These conjugates inhibit HIV-1 protein synthesis and also strongly inhibit HIV-1 production in activated CD4⁺ T cells from infected patients.

We purified pokeweed antiviral protein (PAP) to homogeneity from pokeweed leaves (*Phytolacca americana*)¹², and tested its effects on HIV-1 production in primary human CD4⁺ cells. The 50% inhibitory dose (ID₅₀) for HIV-1 p24 production, in both primary CD4⁺ T cells and macrophages, was $\sim 5 \times 10^3$ pM PAP (Fig. 1). By contrast, a 10-fold greater concentration did not inhibit proliferation of uninfected phytohaemagglutinin (PHA)-stimulated CD4⁺ T cells or bone marrow progenitor cells (Fig. 1). PAP's antiviral effect thus extends to a retrovirus, HIV-1.

We next asked whether PAP targeted to CD4⁺ T cells, by conjugation with monoclonal antibodies (mAb) reactive with normal antigens on CD4⁺ cells, would be more effective at inhibiting HIV-1 replication. Internalization of PAP-mAb conjugates by mAb induced receptor-mediated endocytosis results in increased delivery of PAP through the plasma membrane compared with nonspecific uptake at high PAP concentrations¹³. Also, PAP-mAb conjugates have a plasma half-life of 16-18 h compared with <30 min for free PAP¹⁴ (F.U. and G. Fuller, manuscript in preparation), and thus PAP conjugates could have increased therapeutic potential. We covalently linked PAP to mAb reactive with CD5, CD7 and CD4 antigens, expressed on CD4⁺ T cells, as previously described^{15,16}. The anti-CD4 mAb used reacts with the CD4 gp120-binding site¹⁷. The purity and composition of the conjugates are shown in Fig. 2. The ID₅₀ values for HIV-1 p24 production were 50 pM for PAP-anti-CD5, 6 pM for PAP-anti-CD7 and 1 pM for PAP-anti-CD4. By contrast, PAP conjugated to anti-CD19, which reacts with B cells but not CD4⁺ T cells, caused only 20% inhibition of p24 production even at 1×10^3 pM (Fig. 1). PAP-anti-CD4 and PAP-anti-CD7 were $\sim 1,000$ times more potent than nonconjugated PAP in inhibiting HIV-1 replication. Only at concentrations at least three orders of magnitude higher than the ID₅₀ for p24 production did PAP-anti-CD4 cause 50% inhibition of proliferation of PHA-stimulated uninfected CD4⁺ T cells or colony formation by bone marrow progenitor cells (Fig. 1). Furthermore, PAP-anti-CD4, at 50 times the ID₅₀ for HIV replication, bound only 3% of the cell-surface CD4 molecules and did not inhibit several CD4-dependent responses (Fig. 1).

Regarding the mechanism whereby PAP inhibits HIV-1 replication, it is unlikely to inhibit HIV-1 reverse transcriptase, as does 3'-azido-3'-deoxythymidine (AZT), a treatment for

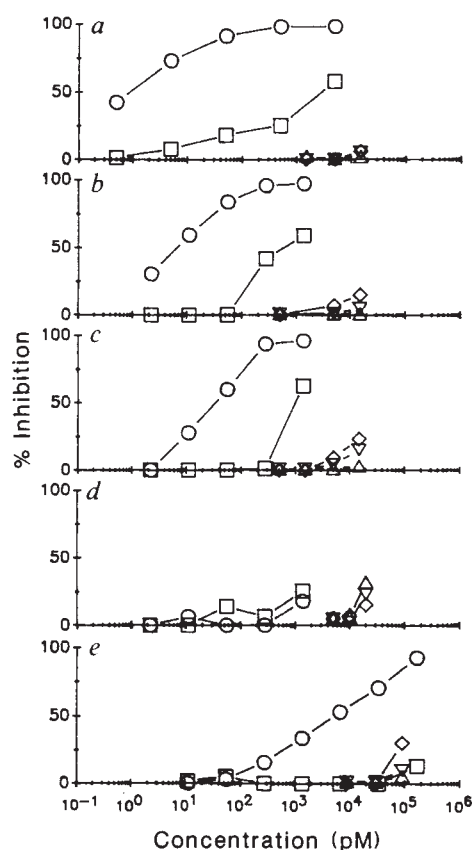


FIG. 1 Inhibition of HIV-1 p24 production in CD4⁺ T cells by PAP and PAP-mAb conjugates at concentrations which do not inhibit proliferation of CD4⁺ T cells or haematopoietic colony formation.

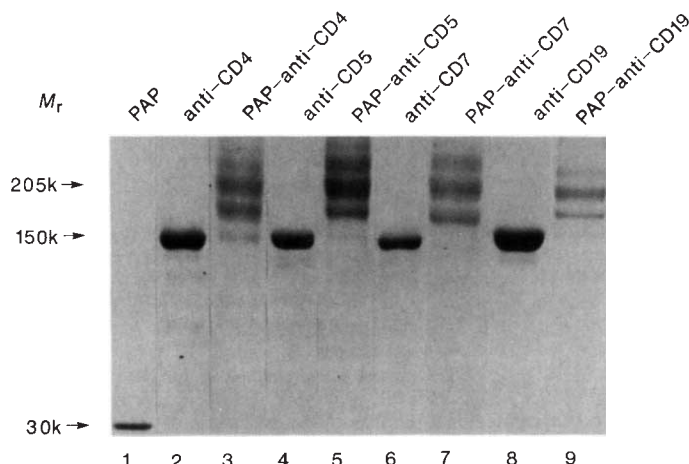
METHODS. PAP was purified by ion-exchange chromatography from spring leaves of pokeweed plants¹² and PAP-mAb conjugates were prepared and purified as previously described¹⁵⁻¹⁶ and as shown in Fig. 2. CD4⁺ T cells were isolated from peripheral blood mononuclear cells by depleting monocytes by adherence to plastic for 1 h at 37 °C, followed by treating for 1 h at 4 °C with mAb to CD8, CD16 and CD20 for 1 h at 37 °C with rabbit complement. Resulting cells, which were >95% viable and at least 90% CD4⁺, were incubated overnight with PHA (3 µg ml⁻¹), then treated for 4 h at 37 °C with PAP or PAP-mAb conjugates. *a*, PAP-anti-CD4; *b*, PAP-anti-CD7; *c*, PAP-anti-CD5; *d*, PAP-anti-CD19; *e*, unconjugated PAP. They were infected for 90 min with LAV-1_{BRU} isolate of HIV-1 at a multiplicity of infection (MOI) of 0.5. After washing off unbound virus, the cells were cultured with PHA, interleukin-2 (IL-2, 4 U ml⁻¹, Biotest) and the designated concentrations of PAP or PAP-mAb conjugates. On day 5, [³H]thymidine ([³H]TdR) incorporated during 6 h into uninfected CD4⁺ cells was determined and on day 7, HIV-1 p24 core antigen concentrations in supernatants were measured by a p24 antigen capture ELISA (Genetic Systems). The mean p24 concentration in supernatants from untreated cells was 1.14 µg ml⁻¹. Toxicity of PAP and PAP-mAb conjugates against normal monopotent myeloid (CFU-GM, colony-forming unit granulocyte macrophage), erythroid (BFU-E, burst-forming unit erythroid), and pluripotent multilineage (CFU-GEMM, colony-forming unit granulocyte-erythroid-macrophage-megakaryocyte) bone marrow progenitor cells was evaluated by colony assays as described³³. Results shown are the mean values for per cent reduction using the formula: % reduction = 100 × 1 - (mean values for treated cells/mean values for untreated cells). ∇, BFC-E; △, CFU-GM; ◇, GEMM; ○, p24 concentration; □, [³H]TdR incorporation. Concentrations of PAP-anti-CD4 of at least 50 pM, which inhibited virus production nearly 100%, did not inhibit CD4-dependent cell functions including (1) antigen specific proliferation of an HLA class II antigen-specific CD4⁺ T-cell clone, (2) generation of cytotoxic T cells in mixed leukocyte cultures or (3) CD4-dependent transmembrane cell signalling measured by mobilization of cytoplasmic Ca²⁺; and 50 pM PAP-anti-CD4 saturated only 3% of cell-surface CD4 molecules (for data see Supplementary Information).

AIDS^{18,19}, because PAP also inhibits replication of viruses lacking this enzyme⁶⁻¹¹. Furthermore, because delaying the onset of treatment of CD4⁺ T cells with PAP or PAP-mAb conjugates until 24 h post-infection did not decrease their efficacy, PAP seems not to inhibit HIV-1 replication at an early step in infection. Additionally, PAP does not inhibit the process of HIV-1 release from cells, as does interferon²⁰, as treatment resulted in a reduction in intracellular HIV-1 protein levels as determined by immunoprecipitation and SDS-PAGE (Fig. 3). In CD4⁺ T

cells treated with 5 pM or 50 pM PAP-anti-CD4, we detected approximately 50% or 90%, respectively, less intracellular gp160 env and p55 and p24 gag proteins 48 h after infection when viral proteins were first detected in untreated infected cells. Our finding that PAP inhibits HIV-1 protein expression is similar to that reported for other viruses^{9,10}. Although PAP can inactivate the 60S subunit of ribosomes and prevent peptide elongation *in vitro*²¹, inhibition of HIV-1 antigen production in infected PHA-stimulated CD4⁺ cells occurred at concentrations of PAP and

FIG. 2 Composition and purity of PAP-mAb conjugates.

METHODS. PAP-mAb conjugates were prepared as previously described^{15,16}. The monoclonal antibodies used were T101³⁴ (IgG2a, anti-CD5/Tp67), G3.7³⁵ (IgG1, anti-CD7/Tp41), B43³⁶ (IgG₁, anti-CD19/Bp95) and G17-2³⁷ (IgG₁, anti-CD4). The chemical composition and purity of the PAP-mAb conjugates were assessed by SDS-PAGE on a 5% running gel followed by gel scanning. Under nonreducing conditions, the conjugates were composed of four *M_r* species of 180K, 210K, 240K and 270K, corresponding to 1:1, 2:1, 3:1 and 4:1 molar PAP:mAb ratios. No unconjugated PAP was detected and unconjugated monoclonal antibody contamination was 2.0% in PAP-anti-CD4, 2.4% in PAP-anti-CD7, 3.0% in PAP-anti-CD5 and 3.5% in PAP-anti-CD19 preparations. Under reducing conditions, only three bands were seen corresponding to heavy and light chains of the antibodies, and PAP (not shown); hence, the conjugates used in this study were >96% pure. For calculations of molar concentrations, an average *M_r* of 210K was used for all PAP-mAb conjugates because the conjugates consisted primarily of 2:1 molar PAP-mAb ratios. PAP-mAb conjugates bound to target cells bearing the respective antigens with the same affinity and specificity as unconjugated monoclonal antibody and their binding could be selectively blocked by 100-fold molar excess of respective unconjugated monoclonal antibody, as determined by immunofluorescence and flow cytometry (data not shown). The ribosome-inhibitory activity of purified conjugates and unconjugated PAP was tested in a cell-free translation system from rabbit reticulocytes³⁸. The IC₅₀ values were 25 pM for PAP-anti-CD4, 34 pM for PAP-anti-CD5, 39 pM for PAP-anti-CD7, 21 pM for PAP-anti-CD19, and 7.7 pM for unconjugated PAP.



PAP-anti-CD4 well below those that inhibited proliferation of normal PHA-stimulated CD4⁺ cells (Fig. 1). Cell membrane changes associated with HIV-1 infection²² possibly result in increased intracellular levels of PAP and PAP-mAb conjugates.

At least 1 in 100 CD4⁺ T cells of HIV-1-infected patients can contain proviral DNA^{3,4}, but most of these cells are latently infected and do not produce HIV-1 or express HIV-1 proteins²³. We have found that stimulation of patients' peripheral blood lymphocytes (PBL) with an IgG₁ monoclonal antibody to CD3 (G19-4)²⁴ is very effective at inducing production of HIV-1 (J.M.Z., P.A.M., J.S. and J.A.L., manuscript in preparation). Although CD4 is downregulated by HIV-1 infection of CD4⁺ T cells *in vitro*^{25,26}, patients' latently infected cells express CD4^{3,4} and so we asked if PAP-anti-CD4 would inhibit HIV-1 production in patients' CD4⁺ cells activated with anti-CD3. PAP-anti-CD4, which is the most specific at targeting PAP to CD4⁺ cells, was found to inhibit HIV-1 production in patients' PBL virtually completely at a concentration as low as 0.5 pM PAP-anti-CD4, which did not inhibit proliferation of the stimulated PBL (Fig. 4a). Notably, treatment of patients' anti-CD3-stimulated PBL with 5 pM PAP-anti-CD4 inhibited production of HIV-1 for at least 22 days, even when the cells were washed free of the conjugate on day 5 (Fig. 4b). The possibility that treatment with PAP-anti-CD4 may have eliminated most of the activated

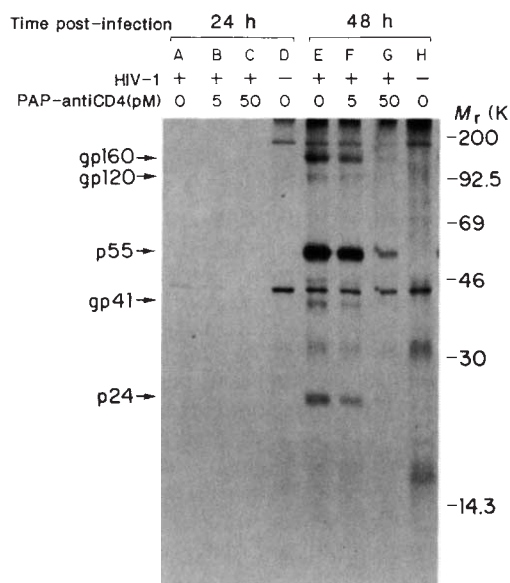


FIG. 3 Inhibition of HIV-1 protein expression in CD4⁺ T cells by PAP-anti-CD4. Radioimmunoprecipitation (RIP) of cell lysates from HIV-1 infected CD4⁺ PBL treated with 0 pM (lanes A and E), 5 pM (lanes B and F), or 50 pM (lanes C and G) PAP-anti-CD4. Lanes D and H correspond to control RIPs of cell lysates from uninfected normal PBL. Samples were assayed at 24 h (lanes A–D) and 48 h (lanes E–H) after infection.

METHODS. CD4⁺ T cells were isolated from blood of a seronegative donor, as described in Fig. 1, and were pretreated overnight with 5.0 or 50 pM PAP-anti-CD4. The cells were infected for 1 h with HIV-1 (MOI 10) and cell-free HIV-1 was washed off cells. The cells were cultured with or without PAP-anti-CD4 and with IL-2 (4 U ml⁻¹) and PHA (3 µg ml⁻¹). At 24 and 48 h post-infection, the cells were pulse radiolabelled with [³⁵S]methionine and [³⁵S]cysteine (100 µCi ml⁻¹) for 2 h then washed, collected, and lysed in RIP lysis buffer (1% NP40, 0.5% DOC and 0.1% SDS in PBS). The samples were immunoprecipitated using polyclonal anti-HIV human sera (Trimar) as described previously³⁹, and fractionated by SDS-PAGE in 11.5% acrylamide gels. These concentrations of PAP-anti-CD4 did not inhibit proliferation of CD4⁺ T cells (Fig. 2) or decrease total cellular protein synthesis as determined by incorporation of [³⁵S]methionine and [³⁵S]cysteine TCA-precipitable radioactivity in equivalent volumes of cell lysates.

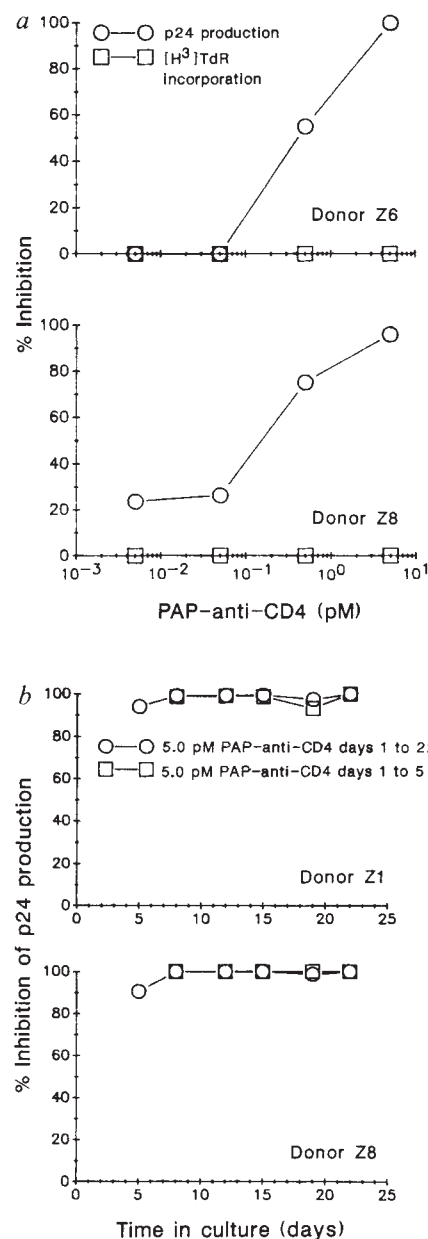


FIG. 4 Inhibition of HIV-1 p24 production in activated PBL from HIV-1 seropositive donors by treatment with PAP-anti-CD4.

METHODS. a, PBL from 2 asymptomatic, HIV-1 seropositive donors were cultured with an IgG₁ mAb to CD3 (G19-4, 1 µg ml⁻¹) and IL-2 (Electronucleonics), to activate T-cell proliferation and HIV-1 replication, in the presence of various concentrations of PAP-anti-CD4 conjugate. On day 5 after activation, [³H]TdR incorporation into cells was determined and on day 7, p24 concentrations in supernatants were measured. The concentrations of p24 in supernatants of PBL, not treated with PAP-anti-CD4, from donors Z6 and Z8, were 0.4 and 6.6 µg ml⁻¹, respectively. b, PBL from two asymptomatic, seropositive donors were cultured with mAb G19-4 to CD3, with IL-2 and with 5.0 pM PAP-anti-CD4. On day 5, the PAP-anti-CD4 treated cells were washed three times to remove conjugate from the medium. The efficacy of the cell washing procedure was demonstrated by western blot analysis of supernatants from PAP-anti-CD4 treated cells. Using an antibody to PAP, we found that washing the cells three times removed at least 99.5% of the conjugate and this was confirmed using [¹²⁵I]-labelled PAP-anti-CD4 (for data see Supplementary Information). The washed cells were recultured with or without 5.0 pM PAP-anti-CD4 and to induce continued cell proliferation were passaged weekly in medium containing IL-2 into wells to which mAb G19-4 was immobilized. Concentrations of p24 were measured at the indicated time points after stimulation of PBL. Peak concentrations of 21.3 and 25.2 ng ml⁻¹ of p24 were detected on day 12 in supernatants activated from untreated PBL from donors Z1 and Z8. The per cent inhibition of p24 production by treatment with PAP-anti-CD4 was calculated as in Fig. 1.

latently infected cells was supported by our finding that treatment of patients' PBL with 10 pM PAP-anti-CD4 reduced the frequency of infected cells by at least 97% as determined by limiting dilution-cocultivation with normal PHA blasts in the absence of the conjugate.

Our approach of targeting the antiviral agent, PAP, to uninfected or latently infected CD4⁺ cells, by conjugation with monoclonal antibodies reactive with normal antigens on CD4⁺ cells differs from other new approaches which depend upon expression of HIV-1 envelope antigens, gp41 or gp120, on infected cells²⁷⁻³⁰. For example, ricin A chain (RTA) or pseudomonas exotoxin conjugated to soluble CD4 can kill gp120-expressing cells^{27,28}. Similarly, monoclonal antibodies against gp120 or gp41 conjugated to RTA can kill HIV-1 infected cells^{29,30} late in the infectious cycle when some mature virus may have already been produced. PAP conjugated to monoclonal antibody against normal antigens on CD4⁺ cells also lacks certain potential clinical problems associated with conjugates of RTA and antiviral monoclonal antibodies, such as variability in envelope antigens among HIV-1 isolates^{31,32} and the ability of patients' plasma anti-envelope antibodies to block the RTA-antibody conjugates from killing infected cells^{29,30}.

Our results show that PAP inhibits HIV-1 replication in primary CD4⁺ cells and that PAP targeted to CD4⁺ T cells by conjugation with monoclonal antibodies against normal antigens on CD4⁺ cells is uniquely active at picomolar concentrations; the PAP-mAb conjugates are up to 1,000 times more potent than nonconjugated PAP. Furthermore, PAP-anti-CD4 was found to be very effective at inhibiting HIV-1 production for at least several weeks in patients' activated replicating CD4⁺ T cells. These findings, together with observations that PAP-mAb conjugates have an *in vivo* plasma half-life at least 30 times that of nonconjugated PAP¹⁴, suggest that PAP conjugated to monoclonal antibodies against normal CD4⁺ cell-associated antigens may be a useful therapy in HIV-infected humans. Furthermore, because PAP has antiviral activity against several other human viruses⁹⁻¹¹, PAP-mAb conjugates may have clinical potential for treating other viral diseases. □

Received 3 April; accepted 2 July 1990.

- Littman, D. R. *A. Rev. Immun.* **5**, 561-587 (1987).
- Dagliesh, A. G. *et al. Nature* **312**, 763-767 (1984).
- Schnittman, S. M. *et al. Science* **245**, 305-308 (1989).
- Psallidopoulos, M. C. *et al. J. Virol.* **63**, 4626-4631 (1989).
- Irvin, J. D. *Pharmac. Ther.* **21**, 371-375 (1983).
- Duggar, B. M. & Armstrong, J. K. *Ann. Mo. Bot. Gard.* **12**, 359-364 (1925).
- Wyatt, S. D. & Shepherd, R. J. *Phytopathology* **59**, 1787-1792 (1969).
- Tomlinson, J. A., Walker, V. M., Flewett, T. H. & Barclay, G. R. *J. gen. Virol.* **22**, 225-230 (1974).
- Aron, G. M. & Irvin, J. D. *Antimicrob. Ag. Chemother.* **17**, 1032-1036 (1980).
- Teltow, G. J., Irvin, J. D. & Aron, G. M. *Antimicrob. Ag. Chemother.* **23**, 390-395 (1983).
- Ussery, M. A., Irvin, J. D. & Hardesty, B. *Ann. N.Y. Acad. Sci.* **284**, 431-437 (1977).
- Ukun, F. M. *et al. in Autologous Bone Marrow Transplantation* (eds Dicke, K.A., Spitzer, G., Jagannath, S. & Evinger-Hodges, M. J.) 171-178 (Anderson, Houston, Texas, 1989).
- Ukun, F. M. *et al. in Membrane-mediated Cytotoxicity, UCLA Symposium on Molecular and Cellular Biology, New Series* (eds Gale, R. P. & Champlin, R.) **45**, 243-256 (Liss, New York, 1987).
- Ramakrishnan, S. & Houston, L. L. *Cancer Res.* **45**, 2031-2036 (1985).
- Ukun, F. M. *et al. J. exp. Med.* **163**, 347-368, 1986.
- Ukun, F. M. *et al. Antibody, Immunocon. Radiopharm.* **1**, 247-262 (1988).
- Linsley, P. S., Ledbetter, J. A., Kinney-Thomas, E. & Hu, S.-L. *J. Virol.* **62**, 3695-3702 (1988).
- Yarchoan, R. *et al. Lancet* **i**, 132 (1987).
- Fischl, M. A. *et al. New Engl. J. Med.* **317**, 185-191 (1987).
- Poli, G., Orenstein, J. M., Kinter, A., Folks, M. & Fauci, A. S. *Science* **244**, 575-577 (1989).
- Obrig, T. G., Irvin, J. D. & Hardesty, B. *Arch. Biochem. Biophys.* **155**, 278-283 (1973).
- Garry, R. G. *et al. Biosci. Rpts* **8**, 35-39 (1988).
- Harper, M. E., Marselle, L. M., Gallo, R. C. & Wong-Staal, F. *Proc. natn. Acad. Sci. U.S.A.* **83**, 772-776 (1986).
- Hoxie, J. A. *et al. Science* **234**, 1123-1127 (1986).
- McDougal, J. S. *et al. J. Immun.* **135**, 361-363 (1985).
- Ledbetter, J. A. *et al. Eur. J. Immun.* **18**, 525-532 (1988).
- Till, M. A. *et al. Science* **242**, 1166-1168 (1988).
- Chaudhary, V. K. *et al. Nature* **335**, 369-372 (1988).
- Till, M. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Pincus, S. H., Wehrly, K. & Chesebro, B. *J. Immun.* **142**, 3070-3075 (1989).
- Starich, B. R. *et al. Cell* **45**, 637 (1986).
- Weiss, R. A. *et al. Nature* **324**, 572 (1986).
- Ukun, F. M. *et al. J. Immun.* **135**, 3817-3822 (1985).
- Royston, I. *et al. J. Immun.* **125**, 725-728 (1980).
- Ukun, F. M. *et al. J. Immun.* **140**, 2103-2111 (1985).
- Ukun, F. M. *et al. Cancer Res.* **45**, 69-75 (1988).
- Ledbetter, J. A., Fell, H. P., Grosmaire, L. S., Norris, N. A. & Tsu, T. *Molec. Immun.* **24**, 1255-1261 (1987).

- Myers, D. E. *et al. J. Immun. Meth.* **121**, 129-142 (1989).
- Haffar, O., Dowbenko, D. & Berman, P. J. *Cell Biol.* **107**, 1677-1687 (1988).

SUPPLEMENTARY INFORMATION. Requests should be addressed to the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. This work was supported by Oncogen, NIH and the Veterans Administration. F. M. Ukun is a Scholar of the Leukemia Society of America. We thank J. D. Irvin for his help in developing procedures for purifying PAP and PAP-mAb conjugates, L. Terry, M. Langlie and L. Tuel for technical assistance and C. Hagen for preparation of this manuscript.

Effect of ATP on actin filament stiffness

Paul A. Janmey*||, Søren Hvidt†, George F. Oster‡, Jennifer Lamb*, Thomas P. Stossel* & John H. Hartwig*§

* Hematology Unit, Massachusetts General Hospital, Departments of Medicine, || Biological Chemistry and Molecular Pharmacology, and § Anatomy and Cellular Biology, Harvard Medical School, Boston, Massachusetts 02114 USA

† Department of Chemistry, Risø National Laboratory and Roskilde University, DK-4000, Roskilde, Denmark

‡ Departments of Entomology and Biophysics, University of California, Berkeley, California 94720, USA

ACTIN is an adenine nucleotide-binding protein and an ATPase¹. The bound adenine nucleotide stabilizes the protein against denaturation² and the ATPase activity, although not required for actin polymerization, affects the kinetics of this assembly (see ref. 3 for review). Here we provide evidence for another effect of adenine nucleotides. We find that actin filaments made from ATP-containing monomers, the ATPase activity of which hydrolyses ATP to ADP following polymerization, are stiff rods, whereas filaments prepared from ADP-monomers are flexible. ATP exchanges with ADP in such filaments and stiffens them. Because both kinds of actin filaments contain mainly ADP, we suggest the alignment of actin monomers in filaments that have bound and hydrolysed ATP traps them conformationally and stores elastic energy. This energy would be available for release by actin-binding proteins that transduce force or sever actin filaments. These data support earlier proposals that actin is not merely a passive cable, but has an active mechanochemical role in cell function⁴⁻⁶.

The effect of monomer-bound ADP or ATP on filamentous (F-)actin structure is evident in electron micrographs of F-actin assembled from monomers with either ADP or ATP bound (hereafter called ATP- or ADP-actin depending on the nucleotide bound to the free monomer). Figure 1a shows negatively stained ATP-F-actin. The width of negatively stained ATP filaments is 7-8 nm, and they are long and straight, bending only with gentle curvatures. ADP-filaments, however, have irregular diameters often as large as 12 nm in places, and they change direction abruptly at acute angles making them appear convoluted and entangled along their length (Fig. 1b). Figures 1c and d show that the differences in morphology between ATP- and ADP-filaments are also evident in rapidly frozen specimens. Mg²⁺-paracrystals formed from ATP- and ADP-filaments also show structural differences (Fig. 1e, f). Filaments in ATP-actin paracrystals are generally parallel and have smooth edges, but paracrystals of ADP-actin have ragged borders because of poorly aligned fibres. Differences in the arrangement of subunits within ADP- and ATP-filaments are apparent in high resolution platinum replicas of the filaments (Fig. 1g, h). ATP-filaments have diameters of 11.5 nm and monomers are visible crossing over every 37 nm; ADP-filaments are considerably wider, and crossovers are difficult to find. Irreversible denaturation was not responsible for the morphology of ADP-filaments, because addition of ATP to ADP-monomer solutions yielded filaments identical to those in Fig. 1a.

Figure 2 shows that ADP-filaments rapidly straighten when

diluted into ATP-containing solutions, and rapid exchange of ATP into the ADP-filaments was documented using the ATP analog etheno-ATP which increases fluorescence upon binding actin⁷ (Fig. 2c). The filaments used in these experiments were capped by gelsolin which inhibits actin monomer exchange with the fast-exchanging ends of F-actin⁸. Formation of ATP-filaments *de novo* from actin monomers cycled from the slow exchanging ends of gelsolin-capped ADP filaments is too slow to occur within this time^{9,10}. Etheno-ATP (ϵ -ATP) binding is an apparent first order process with a rate constant of 0.0010 s^{-1} , near the value which has been measured for dissociation of ADP from ADP-G-actin⁷. Moreover, the rate does not depend on the number of filament ends, which was varied by inclusion of gelsolin at molar ratios from 1:34 to 1:500 (ref. 9). Therefore, monomer exchange, which occurs only at filament ends, is not required for the measured ϵ -ATP binding, and the mechanism by which actin filaments straighten is most likely an exchange of ATP with ADP bound to actin subunits in the interior of filaments which has been inferred by others¹¹⁻¹³. In contrast to ADP-actin filaments, ATP-F-actin bound ϵ -ATP slowly, suggesting that the nucleotide, whether ATP, ADP, or ADP-P_i, was bound very tightly or kinetically trapped in ATP filaments¹⁴.

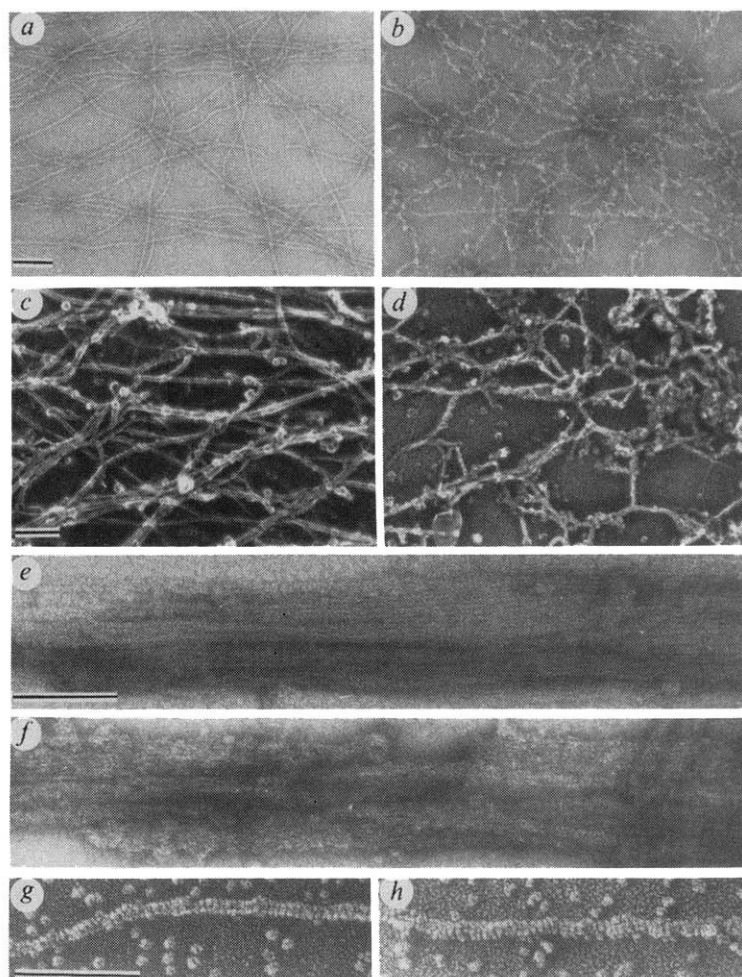
The ultrastructural differences between ADP- and ATP-filaments predict that they should exhibit different mechanical properties. Solutions of long, overlapping rigid rodlike polymers subjected to oscillatory shear display a relatively large dynamic storage modulus (G'), that is, the elastic resistance to deformation (strain) by the shear forces, and a relatively low value of the loss modulus (G''), which reflects the energy dissipated by viscous resistance. Moreover, as long as the filaments are long

enough to impede each others' rotational diffusion, G' is independent of the frequency of oscillation over a range of frequencies greater than the reciprocal of the rotational relaxation time, and G'' is small, as stiff rods have no modes of motion other than translational diffusion, which is also very slow, by which to dissipate mechanical energy^{15,16}. Figure 3a and b shows that the rheology of ATP-actin fulfils these criteria and is therefore entirely consistent with ATP-filaments being long stiff rods.

Flexible filaments are more entangled than stiff rods, and the entanglements (Fig. 1b, d) act as temporary crosslinks which resist deformation in response to high but not low frequency oscillations^{17,18}. Flexible filaments can also have higher values of G' and G'' than rigid rods of the same length because the latter contribute only one motion per polymer, whereas flexible chains undergo a spectrum of motions¹⁹, and shear moduli at a given frequency depend on all viscoelastically active motions with characteristic times longer than the reciprocal of that frequency. Both G' and G'' of ADP-actin display frequency dependence, suggesting the presence of molecular motions on a time scale of 0.01–100 s which are absent from ATP-actin, and which would result from filament flexibility. Adding ATP to ADP-actin monomers before their polymerization caused the viscoelastic properties to recover to those of ATP actin (Fig. 3a, b) as did the addition of tropomyosin, which laminates F-actin by binding to its helix groove, increasing the stiffness of actin filaments²⁰ and inhibiting filament breakage²¹, thereby confirming that the rheologic differences between ATP- and ADP-filaments are related to differences in stiffness.

A second prediction for the rheology of rodlike filaments is

FIG. 1 Ultrastructure of F-actin assembled from either ATP- (a, c, e, g) or ADP-monomeric actin (b, d, f, h). Actin was purified from rabbit skeletal muscle. ADP- and ATP-actin were prepared from the same starting actin solution in 0.2 M CaCl₂, 0.5 mM ATP, 0.5 mM 2-mercaptoethanol, and 2 mM Tris, pH 7.0, (buffer A) and polymerized at concentrations of 24 μ M with either 2 mM MgCl₂ and 100 mM KCl (a, b, e, f) or 2 mM MgCl₂ alone (c, d, g, h). ADP-actin was prepared from the ATP actin by the addition of 1 mM glucose and hexokinase according to the method of Pollard³³ and incubated for 2 h at 25 °C. Alterations in the critical actin monomer concentrations, indicative that the samples contained ATP- or ADP-actin, were documented by a fluorometric assay³⁴. ATP-actin solutions contained the same amount of glucose but lacked hexokinase. a, 2.4 μ M ATP-actin after negative staining with 1% uranyl acetate. b, 2.4 μ M ADP-actin after negative staining as in a. c, Replica of 12 μ M ATP-actin in 2 mM MgCl₂ prepared by rapid freezing on a helium-cooled copper block (Med-Vac, Inc), fracturing with a liquid-nitrogen cooled knife (Cressington CFE-50) at -120 °C, freeze-drying for 18 min at -100 °C, and coating with platinum at 45° and carbon at 90°. d, 12 μ M ADP-actin prepared as in c. e, Structure of negatively stained ATP-actin paracrystals (48 μ M) prepared by dialysis versus 50 mM MgCl₂. f, Paracrystals of ADP-actin after negative staining as in e. g and h, Comparison of the substructure of ATP- (g) and ADP-filaments (h). 2 μ l of 7.2 μ M F-actin, polymerized at 24 μ M with 2 mM MgCl₂, was placed on mica fragments as described by Heuser³⁵ for 30 s and rapidly frozen. Samples were brought to -100 °C, cleaved with a liquid nitrogen cooled knife and rotary coated with 1.6 nm of platinum at 10° and 4 nm of carbon at 90°. Scale bars, 0.1 μ m.



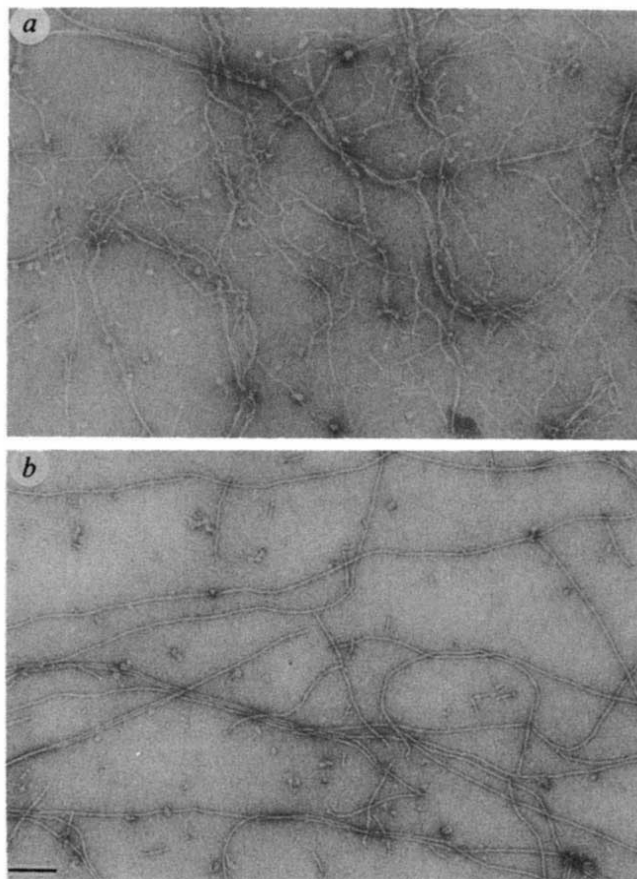
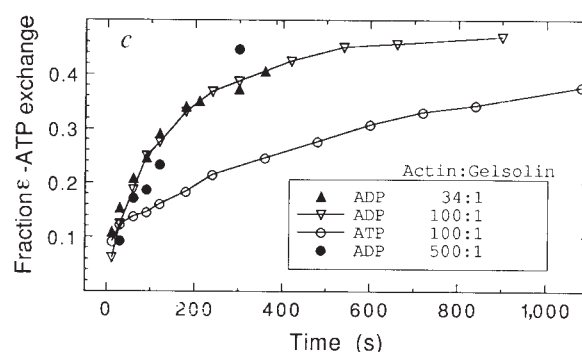


FIG. 2 Rapid conversion of ADP-F-actin to ATP-F-actin. ADP-F-actin ($12 \mu\text{M}$) was prepared for negative staining as described in the legend to Fig. 1, and diluted to $2.4 \mu\text{M}$ in buffer A with 2 mM MgCl_2 (a) or the same buffer incubated with 1 U ml^{-1} hexokinase and 1 mM glucose for 2 h (b) Scale bar, $0.1 \mu\text{m}$. c, Fluorescence increase after adding ADP-F-actin to etheno-ATP (ϵ -ATP): evidence for ATP-ADP exchange. ϵ -ATP was added to ATP- or ADP-actin filaments that had been polymerized in the presence of various amounts of gelsolin, as shown. The fraction of ϵ -ATP bound to actin was calculated from the resulting fluorescence increase⁷ by the relation [fraction bound] = $(F(t) - A)/(F_\infty - A)$ where $F(t)$ is the sample fluorescence at time t , F_∞ is the fluorescence after 20 h , and A the fluorescence of ϵ -ATP in the absence of actin. All samples contained $24 \mu\text{M}$ ϵ -ATP, and $24 \mu\text{M}$ actin in buffer A with 2 mM MgCl_2 except for the 1:500 actin:gelsolin sample which contained $37 \mu\text{M}$ ϵ -ATP and $57 \mu\text{M}$ actin in 2 mM MgCl_2 and 150 mM KCl .



that their elastic storage modulus should increase with increasing strain amplitude (strain hardening)^{15,16,22}. This prediction is also satisfied by ATP filaments (Fig. 3c) for strains below 10%. At larger strains G' shows an abrupt decrease which is due to mechanical breakage of the filaments¹⁶. The elasticity, as exemplified by a low ratio of G'' to G' is also high over the range of strains up to 10% (Fig. 3d). In contrast, ADP-actin shows no evidence of strain-hardening, consistent with the expectation for flexible filaments, and G' decreases at even the lowest measurable range of strains, suggesting that these filaments are even more fragile than ATP-filaments. The ratio G''/G' also increases at much smaller strains for ADP than for ATP-filaments, suggesting the presence of more molecular motions that dissipate energy. The effects of tropomyosin on ADP-actin, causing a recovery of strain-hardening (Fig. 3c) and a lowered G''/G' ratio (Fig. 3d), confirm that ADP-actin alone is highly flexible.

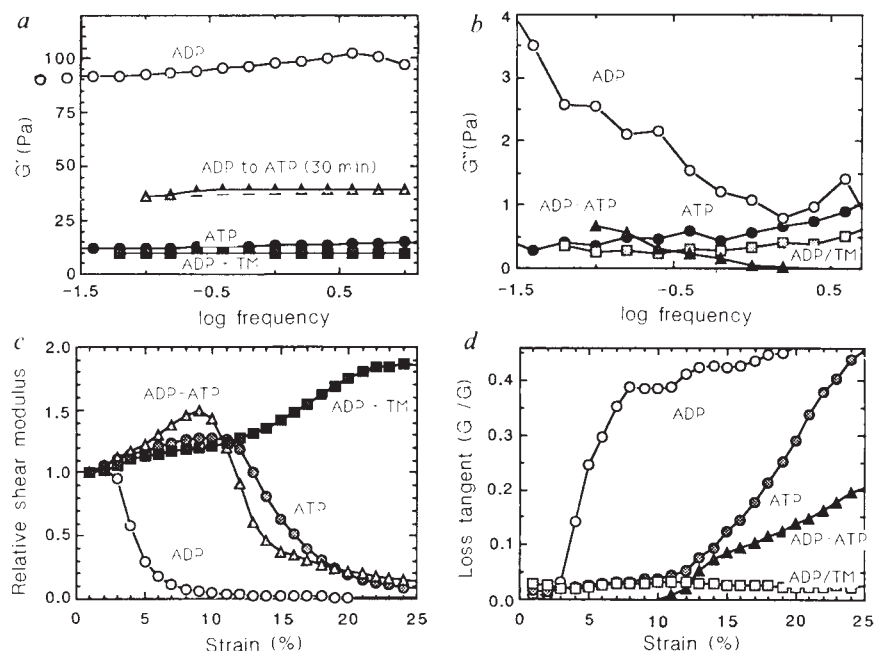
Dynamic light scattering provides additional evidence for a difference between the diffusive motions of ADP- and ATP-actin filaments, because the intensity autocorrelation functions of light scattered from long, overlapping polymers is dominated by intramolecular motions^{20,23,24}. Figure 4 shows the intensity autocorrelation functions of light scattered from equal concentrations of ATP- or ADP-actin. The faster decay observed for ADP-actin suggests that the intramolecular motions of ADP-actin filaments are more rapid than those of ATP-actin. If the increased rate of decay in intensity autocorrelation of ADP filaments is due to their greater flexibility, compared with ATP-filaments, then further stiffening of ATP-filaments with tropomyosin would have the opposite effect on the correlation functions, as shown in Fig. 4, consistent with previous dynamic light scattering studies²⁰.

We have shown that the ultrastructural morphology and viscoelastic properties of actin filaments polymerized from ATP-

or ADP-monomers are different, even though at steady state both types of filament contain primarily actin subunits with bound ADP. As the conformations of ADP and ATP-monomers differ^{6,25,26}, the structures of filaments assembled from ATP- and ADP-monomers may also vary and remain altered after filament-bound ATP hydrolyses to ADP on each filament subunit. A possible basis for the persistence of filament stiffness after ATP hydrolysis is that the conformation of a filament formed from ATP-monomers is trapped in a strained state. When several ATP-bound monomers have added to the end of a filament, multiple contacts with its neighbours constrain a subunit within the filament interior. If the ATP now hydrolyses, these contacts prevent the subunit from converting to the lowest energy configuration adopted by an ADP-monomer adding to a filament without hydrolysis from an ATP-bound subunit. Either of two competing structural models for F-actin, which give primary emphasis to either longitudinal or transverse bonds²⁷⁻²⁹, can account for storage of elastic energy, as long as an appreciable number of both types of bonds is allowed. As ATP is not regenerated by ADP-actin in the presence of high inorganic phosphate ion (P_i) concentrations³, it is not surprising that addition of P_i concentrations from 0.1 mM to 10 mM did not affect the rheological and dynamic light scattering differences between ATP- and ADP-actin (unpublished experiments). The kinetic stability of ATP filaments is confirmed by the finding that ATP-actin maintains its rheologic properties long after free ATP is hydrolysed to ADP (P.A.J., S.H., G.F.O., J.L., T.P.S. and J.H.H., unpublished experiments).

The energy stored in the strained bonds may be utilized during dynamic remodelling of actin filament complexes. The protein gelsolin very efficiently severs F-actin without additional energy input from nucleotide hydrolysis³⁰, even though the free energy required to break a filament, estimated from rates of spontaneous breakage and annealing²⁹, is much greater than that calculated

FIG. 3 Viscoelastic measurements. *a* and *b*, Dynamic storage (G') and loss (G'') shear moduli, respectively, for actin polymerized under different conditions and measured at 1% strain amplitude over a range of frequencies. The shear modulus, a measure of the force required to produce a given deformation, depends on both the frequency of the measurement and the magnitude of deformation, the strain. By imposing a sinusoidally varying strain on the sample and measuring the magnitude and phase of the resulting stress (force per unit area), the dynamic storage (G') and loss (G'') moduli were computed, which quantify the mechanical energy elastically stored or dissipated, respectively, during a cycle of deformation¹⁷. *c*, Strain amplitude dependence of G' measured at $\omega = 1 \text{ rad s}^{-1}$. Measurements were begun at low strains and the maximum strain of the oscillations was increased in successive measurements. *d*, Loss tangent (ratio of G'' to G') as a function of strain for the same samples. For the experiments shown, ADP- and ATP-actin were prepared from the same starting material. A solution of F-actin (polymerized in the presence of ATP) was divided in half and F-actin in each collected by sedimentation. One resultant actin pellet was resuspended into buffer containing 5 mM Tris, 0.2 mM CaCl_2 , 0.2 mM DTT, 0.5 mM ATP, pH 7.4 (ATP-buffer). The other resultant pellet was resuspended in buffer with 0.5 mM ADP in place of ATP (ADP-buffer). The actin suspensions were dialysed at 4 °C against either ATP-buffer or ADP-buffer for 48 h, with three changes of 2 litres of each buffer. Each sample was divided into aliquots which were frozen in liquid nitrogen and thawed 30 min before polymerization which was initiated by the addition of 2 mM MgCl_2 and 150 mM KCl, and which occurred between the plates of the rheometer. The ATP concentrations of the dialysates, determined by the luciferase/luciferin assay (Sigma) showed less than 8 μM ATP, presumably due to the activity of myokinase, contaminating an ADP buffer diluted to 200 μM total nucleotide. For some experiments ADP-actin was prepared from ATP-actin by the glucose/hexokinase method of Pollard³³, and rheologic measurements revealed similar differences between ATP- and ADP-actin. The sample marked ADP to ATP was prepared by adding 2 mM ATP to ADP-actin for 30 min before polymerization. The sample marked ADP + TM was prepared by polymerizing ADP-actin in the presence of 0.7 mg ml^{-1} rabbit skeletal muscle tropomyosin (TM) prepared



by the method of Greaser and Gergeley³⁶. The differences in the rheologic data are not the consequence of large discrepancies in filament length or polymer concentration. The total protein content was adjusted to account for the 3–5 μM difference in critical concentration between ADP- and ATP-actin³, which was confirmed by the light scattering experiments shown in Fig. 4. Filament length was regulated by polymerization of ADP- and ATP-actin in the presence of a 1:1,000 molar ratio of gelsolin to F-actin. As gelsolin reacts with both ADP- and ATP-actin to form stable caps at the barbed ends of filaments³⁰, both samples contain filaments with a number average of 1,000 subunits. The concentrations of ATP- and ADP-actin were 34 μM (1.43 mg ml^{-1}) and 38 μM (1.6 mg ml^{-1}) to allow for the approximately 4 μM difference in critical concentration. All samples contained 34 nM gelsolin. The ADP-actin:TM molar ratio was 7:4. The dynamic shear moduli at 20 °C were measured with a Rheometrics RFS8500 instrument the operation of which is described in ref. 16.

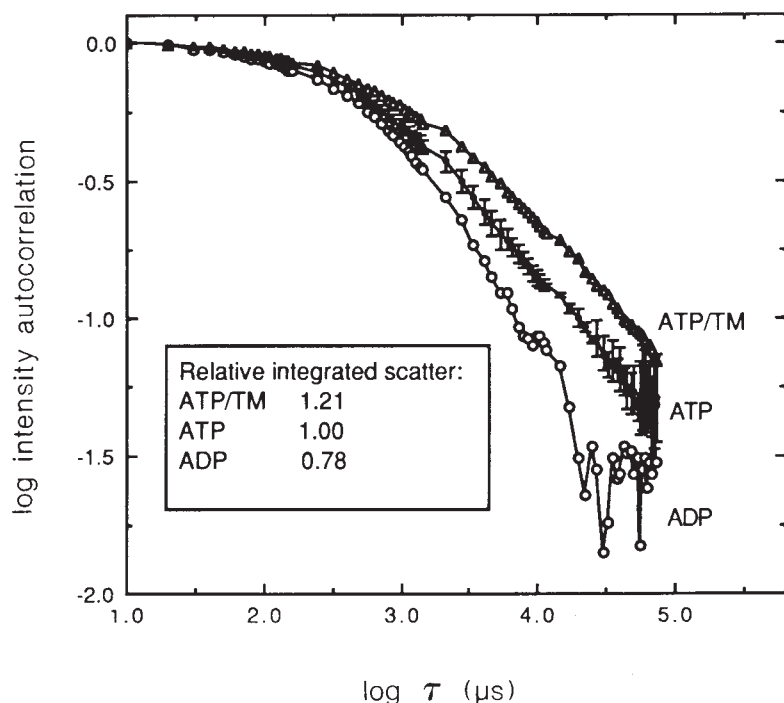


FIG. 4 Dynamic light scattering of F-actin with ADP, ATP and tropomyosin. The time correlation of light scattered from solutions of long, interpenetrated filaments is dominated by intramolecular motions of the polymer chains with a characteristic distance related to the scattering vector ($\sim 40 \text{ nm}$ for these experiments). In particular, the quasielastic light scattering spectra of F-actin solutions reflect internal motions of the filaments^{20,23,24}. Intensity autocorrelation is shown as a function of the delay time τ for ADP-actin, ATP-actin, and ATP-actin polymerized with tropomyosin (TM) at a 4:7 molar ratio to actin, from 633 nm light scattered at 90° degrees using a Brookhaven Instruments BI3020 correlator. A multiple sample time technique was used to increase the time range of the measurements using four simultaneous sample times and light was collected for 30 min for each correlation function. The error bars for ATP-actin represent the difference between measurements of 0.5 mg ml^{-1} and 2.0 mg ml^{-1} samples. The difference in critical monomer concentration is 3 μM , as calculated from the difference in integrated scattering intensities, within the range of published values. The concentration of ADP-actin was 0.5 mg ml^{-1} and that of ATP-actin-TM was 2.0 mg ml^{-1} . Control experiments showed no systematic variation of correlation functions of ATP-actin over this concentration range.

from the binding affinity of gelsolin and other filament capping proteins for F-actin³¹. Actin-myosin interactions may also utilize strain energy stored in the actin filament. The present data are consistent with proposals by others^{4-6,32} that conformational

changes between subunits within the thin filament may participate in the motions that occur during a crossbridge cycle, and that actin filaments are not merely passive cables along which myosin crossbridges tread. □

Received 29 December 1989; accepted 17 July 1990.

1. Straub, F. B. & Feuer, G. *Biochim. biophys. Acta* **4**, 455-470 (1950).
2. Asakura, S. *Arch. Biochem. Biophys.* **92**, 140-151 (1961).
3. Carlier, M.-F. *Int. Rev. Cytol.* **115**, 139-170 (1989).
4. Szent-Gyorgyi, A. & Prior, G. *J. molec. Biol.* **15**, 515-538 (1966).
5. Yanagida, T. & Oosawa, F. *J. molec. Biol.* **126**, 507-524 (1978).
6. Schutt, C. E., Lindberg, U., Myslik, J. & Strauss, N. *J. molec. Biol.* **209**, 735-746 (1989).
7. Neidl, C. & Engel, J. *Eur. J. Biochem.* **101**, 163-169 (1979).
8. Yin, H. L. & Stossel, T. P. *Nature* **281**, 583-586 (1979).
9. Janmey, P. A. & Stossel, T. P. *J. Muscle Res. Cell Motil.* **7**, 446-454 (1986).
10. Pollard, T. D. *J. Cell Biol.* **103**, 2747-2754 (1986).
11. Brenner, S. L. & Korn, E. D. *J. biol. Chem.* **259**, 1441-1446 (1984).
12. Hozumi, T. *J. Biochem.* **104**, 285-288 (1988).
13. Dancker, P. & Fischer, S. *Z. Naturforsch.* **44c**, 698-704 (1989).
14. Martonosi, A., Gouvea, M. A. & Gergely, J. *J. biol. Chem.* **235**, 1700 (1960).
15. Doi, M. & Kuzuu, N. *J. Polym. Sci. Polym. Phys. Ed.* **18**, 409-419 (1980).
16. Janmey, P. A. *et al.* *Biochemistry* **27**, 8218-8227 (1988).
17. Ferry, J. *Viscoelastic properties of polymers* (John Wiley, New York, 1980).
18. deGennes, P. G. *Macromolecules* **9**, 587-593 (1976).
19. Okubo, N., Komatsubara, M., Nakajima, H. & Wada, Y. *Biopolymers* **15**, 929-947 (1976).
20. Newman, J. & Carlson, F. D. *Biophys. J.* **29**, 37-48 (1980).

21. Oosawa, F. & Asakura, S. *J. molec. Biol.* **161**, 217-227 (1982).
22. Janmey, P., Amis, E. & Ferry, J. *J. Rheol.* **27**, 135-153 (1983).
23. Fujime, S. & Ishiwata, S. *J. molec. Biol.* **62**, 251-265 (1971).
24. Schmidt, C. F., Baermann, M., Isenberg, G. & Sackmann, E. *Macromolecules* **22**, 3638-3649 (1989).
25. Rich, S. A. & Estes, J. E. *J. molec. Biol.* **104**, 777-792 (1976).
26. Faulstich, H., Merkle, I., Blackholm, H. & Stournaras, C. *Biochemistry* **23**, 1608-1612 (1984).
27. Aebi, U., Millonig, R., Salvo, H. & Engel, A. *Ann. NY. Acad. Sci.* **483**, 100-119 (1986).
28. Egelman, E. H. *J. Muscle Res. Cell Motil.* **6**, 129-151 (1985).
29. Erickson, H. J. *J. molec. Biol.* **208**, 465-474 (1989).
30. Coue, M. & Korn, E. *J. biol. Chem.* **261**, 3628-3631 (1986).
31. Coue, M. & Korn, E. D. *J. biol. Chem.* **260**, 15033-15041 (1985).
32. Ando, T. *J. Biochem.* **105**, 818-822 (1989).
33. Pollard, T. D. *J. Cell Biol.* **99**, 769-777 (1984).
34. Kouyama, T. & Mihashi, K. *Eur. J. Biochem.* **114**, 33-38 (1981).
35. Heuser, J. *J. molec. Biol.* **169**, 155-195 (1983).
36. Greaser, M. & Gergely, J. *J. biol. Chem.* **246**, 4226-4232 (1971).

ACKNOWLEDGEMENTS. This work was supported in part by the United States Public Health Service, the American Cancer Society, the MacArthur Foundation, the Danish National Science Foundation, the Whitaker Foundation, and NATO. We are grateful to M. Desisto for technical assistance and M. Schliwa for many valuable discussions.

Crystal structure of ovalbumin as a model for the reactive centre of serpins

Penelope E. Stein*, Andrew G. W. Leslie†, John T. Finch†, William G. Turnell†, Paul J. McLaughlin† & Robin W. Carrell*

* Department of Haematology, University of Cambridge, Hills Road, Cambridge CB2 2QH, UK

† MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

THE serpins are a widely distributed family of proteins with diverse functions; they include the key serine protease inhibitors of human plasma as well as noninhibitory homologues such as hormone-binding globulins, angiotensinogen and egg-white ovalbumin¹. Sequence alignment based on the crystal structure of the cleaved form of the archetypal serpin, α_1 -antitrypsin², indicates that the serpins share a common highly ordered structure³. On cleavage of the reactive centre peptide bond, they characteristically undergo a remarkable conformational change, the newly generated C terminus moving some 70 Å to the opposite pole of the molecule. The structure of this post-cleavage form is known, but the conformation of the intact serpins and in particular that of their reactive centre is not. Wright *et al.*'s structure of plakalbumin⁴ (ovalbumin cleaved by subtilisin) has provided evidence for the conformational change that results from cleavage. We have now determined the structure of native ovalbumin to 1.95 Å resolution and have found that the intact peptide loop forming the analogue to the reactive centre of the inhibitory serpins takes the unexpected form of a protruding, isolated helix. This model of the intact structures of the serpins suggests how they may interact with their target proteases.

The serpin reactive centre is believed to act as a pseudosubstrate for the target protease. The inhibitor and protease form a tight complex with potential cleavage at a P_1-P_1' peptide bond (Table 1). The P_1 site primarily defines specificity of inhibition; for example antithrombin, which inhibits serine proteases cleaving at arginyl bonds, has an arginine at P_1 , and antitrypsin, which primarily inhibits elastase, has a methionine. The reactive centre, and the analogous region of the noninhibitory serpins, must lie within an exposed sequence ('the reactive centre loop') of at least 12 residues, as the region $P_{10}-P_2'$ is readily cleaved

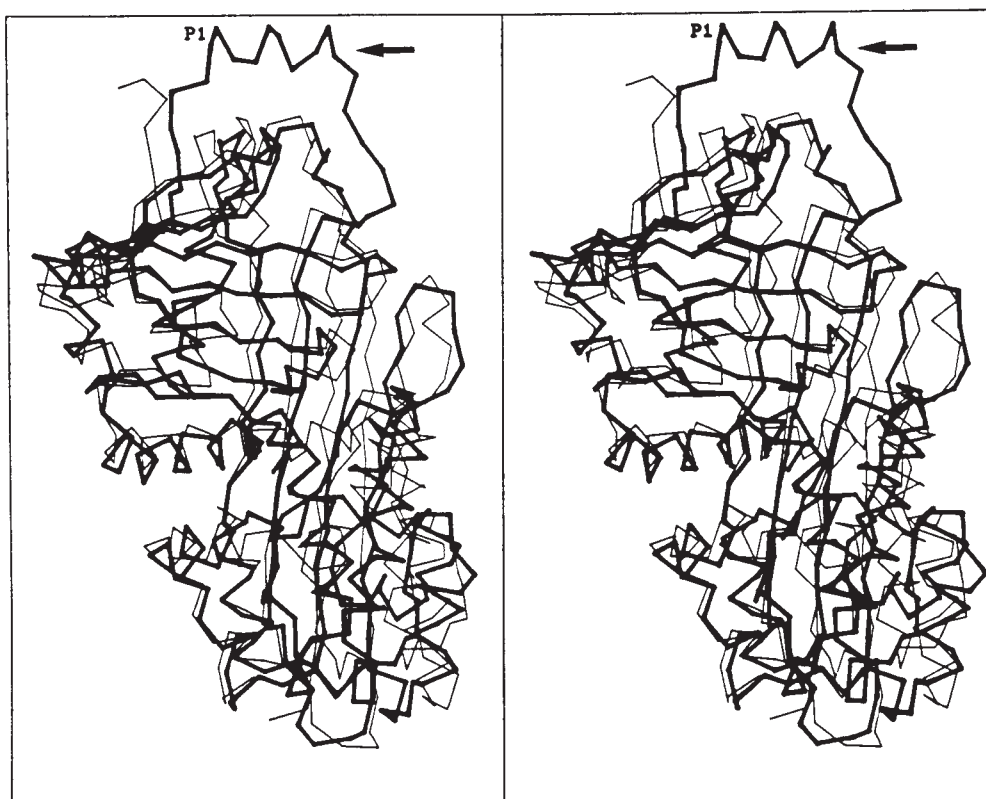
by nontarget proteases. In most serpins this cleavage is accompanied by a conformational change to a more thermostable form^{5,6} in which the cleaved loop moves into the centre of the A sheet to form the fourth strand². Such proteolytically modified serpins are more easily crystallized, and all structures presently solved are in the cleaved form. These structures provide a detailed description of the molecule after the conformational change that occurred on cleavage, but fail to show the structure of the intact reactive centre.

To provide a model for the intact reactive centre we have solved the crystal structure of native ovalbumin. Ovalbumin is a well-conserved member of the serpin family showing sequence identity of 30% with antitrypsin and a putative reactive centre at Ala 358-Ser 359 (residue numbering is based on the sequence alignment with antitrypsin³). On the other hand, it shows no measurable inhibitory activity and does not undergo the typical conformational change on cleavage at this P_1-P_1' site^{7,8}. The recently determined structure of plakalbumin⁴ suggests features that would prevent accommodation of the ovalbumin sequence in the cleaved antitrypsin structure and may cause ovalbumin to be locked in the native conformation. This inherent stability of native ovalbumin makes it a good choice to use as a model of the intact structure for the serpin family as a whole.

The structure determination of intact ovalbumin is described in the legend to Fig. 1, and full details of the structure itself will be reported elsewhere. Here we report the unexpected α -helical form of the reactive centre loop (Figs 1 and 2). This is formed by a 19-residue segment extending from P_{15} Gly to P_2' Glu (Table 1). The P_1-P_1' Ala-Ser is exposed on the final turn of the three-turn helix, which protrudes from the main body of the molecule on two peptide stalks, each of about four residues. Figure 3a, a superposition of three crystallographically independent molecules shows that the helix is mobile; the positions of the helices relative to the protein core vary by 2-3 Å, probably as a result of their different environments in the crystal lattice. The refined atomic temperature factors also illustrate the helical mobility: the average main-chain temperature factor for residues comprising the loop is 28 Å² compared with 22 Å² for the protein as a whole, consistent with a flexible, even wobbly helix that makes few van der Waals contacts with the rest of the molecule (Fig. 2). Despite this evidence of mobility, interpretation of the electron density map for this part of the structure was unambiguous (Fig. 3b).

The structure agrees with known features of native ovalbumin. As expected, the phosphorylated P_9 serine is on the exposed

FIG. 1 A superposition of the α -carbon backbones of ovalbumin (thick bonds) and cleaved antitrypsin (thin bonds) achieved by optimal superposition of helix A, helix B and sheet B in the two structures. The intact reactive centre loop of ovalbumin ($\leftarrow$) and the P₁ site are indicated. The sequence in antitrypsin corresponding to the reactive centre loop of ovalbumin is inserted as an additional strand (A4) in the A sheet as previously reported⁴. Ovalbumin was prepared from single, newly laid hen eggs by ammonium sulphate fractionation and precipitation at pH 4.6¹⁶ followed by DEAE-cellulose ion exchange chromatography at pH 7.0¹⁷ to isolate the major component with two phosphoserines per protein molecule. Ovalbumin crystals were grown by the batch method at 37 °C in 1–3 months. The crystallization medium comprised 25 mg ml⁻¹ ovalbumin, 50% saturated ammonium sulphate and sodium cacodylate buffer (pH 6.4). The crystals were isomorphous with those reported by Miller *et al.*¹⁸, space group P1 ($a=62.9$ Å, $b=84.7$ Å, $c=71.5$ Å, $\alpha=87.5^\circ$, $\beta=104^\circ$, $\gamma=108.5^\circ$) with four ovalbumin molecules in the unit cell. Native ovalbumin data was collected using a synchrotron radiation source and the image plate detector at the EMBL outstation at DESY, Hamburg. A total of 352,293 observations were merged to give 94,362 unique reflections between 30 and 1.95 Å resolution (94% complete) with a merging R -factor of 6.8% on intensities (11.8% at 1.95 Å resolution). The structure was solved by the molecular replacement method using the plakalbumin coordinates⁴ as a starting model. The model was refined using the constrained restrained least-squares minimization program, CORELS¹⁹,



followed by simulated annealing with molecular dynamics using XPLOR²⁰ and then by a restrained parameter least-squares method using PROLSQ²¹ and model building with FRODO²². The current model, which includes 679 ordered water molecules, has a crystallographic R -factor of 16.9% for all data between 6 and 1.95 Å resolution. The root mean squared deviation from ideal values for bond lengths is 0.02 Å and bond angles is 2.9°.

FIG. 2 Stereo view of the reactive centre loop showing the limited nature of the interactions between the helix and the protein core. The axis of the helix runs almost perpendicular to strands C2 and C3 of the C-sheet. The most extensive van der Waals contacts involve residues Ala 353, Asp 356 and Ala 357 of the helix with Ile 224 in the C3 strand. Other interactions include a salt bridge between Asp 356 and Lys 283, although this interaction is only present in two of the four copies of the molecule in the crystallographic unit cell. There is a solvent channel, running between the N-terminal end of the helix and the protein core, which is occupied by ordered solvent molecules in the crystal structure.

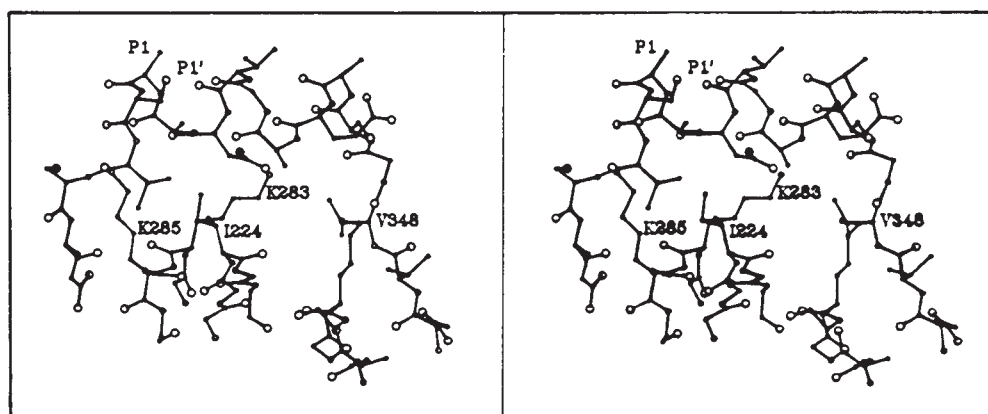


TABLE 1 Sequence alignment of the reactive centre loops of ovalbumin, antitrypsin and antithrombin

Sequence alignment of the reactive centre loops of ovalbumin, antitrypsin and antithrombin																				
	α helix																			
Number (P-P')	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1'	2'	3'	4'	5'
Ovalbumin	G	R	E	V	V	G	S	A	E	A	G	V	D	A	A	S	V	S	—	E
Antitrypsin	G	T	E	A	A	G	A	M	F	L	E	A	I	P	M	S	I	P	P	E
Antithrombin	G	S	E	A	A	A	S	T	A	V	V	I	A	G	R	S	L	N	P	N

face of the helix and the known cleavage sites^{9,10} are all within the helical sequence, although the helix probably unfolds before it is cleaved. Although it is possible that the ovalbumin structure in this region is atypical of the family as a whole, the remarkable similarity between the structures of intact ovalbumin and cleaved antitrypsin in other respects (Fig. 1) indicates that the helix is present in all serpins. The aligned sequences of inhibitory members of the family show no features that preclude such a conformation (Table 1) and in fact a helix was originally strongly predicted for the N-terminal part (P₁₀-P₄) of the antitrypsin reactive centre sequence². We have built a preliminary model of the antitrypsin reactive centre based on the ovalbumin structure, in which the P₈ methionine of antitrypsin is placed in an

external position where, as observed, it would be susceptible to oxidation without affecting the structure of the loop. The bulky phenylalanine at P₇ is also readily accommodated and although the proline at P₂ causes some local distortion, it does not change the overall conformation of the helix.

In the inhibitory serpins, the sequence of the peptide stalk leading from the A sheet to the commencement of the helix (P₁₅-P₉) is well conserved, showing a dominance of residues with small side chains, principally alanines and glycines from P₁₂-P₉, suggesting that the stalk is flexible. In ovalbumin (a noninhibitor) this sequence homology breaks down with an arginine at P₁₄ and valine residues at P_{12,11}. Pathological mutants of human plasma serpins support the idea that the stalk must

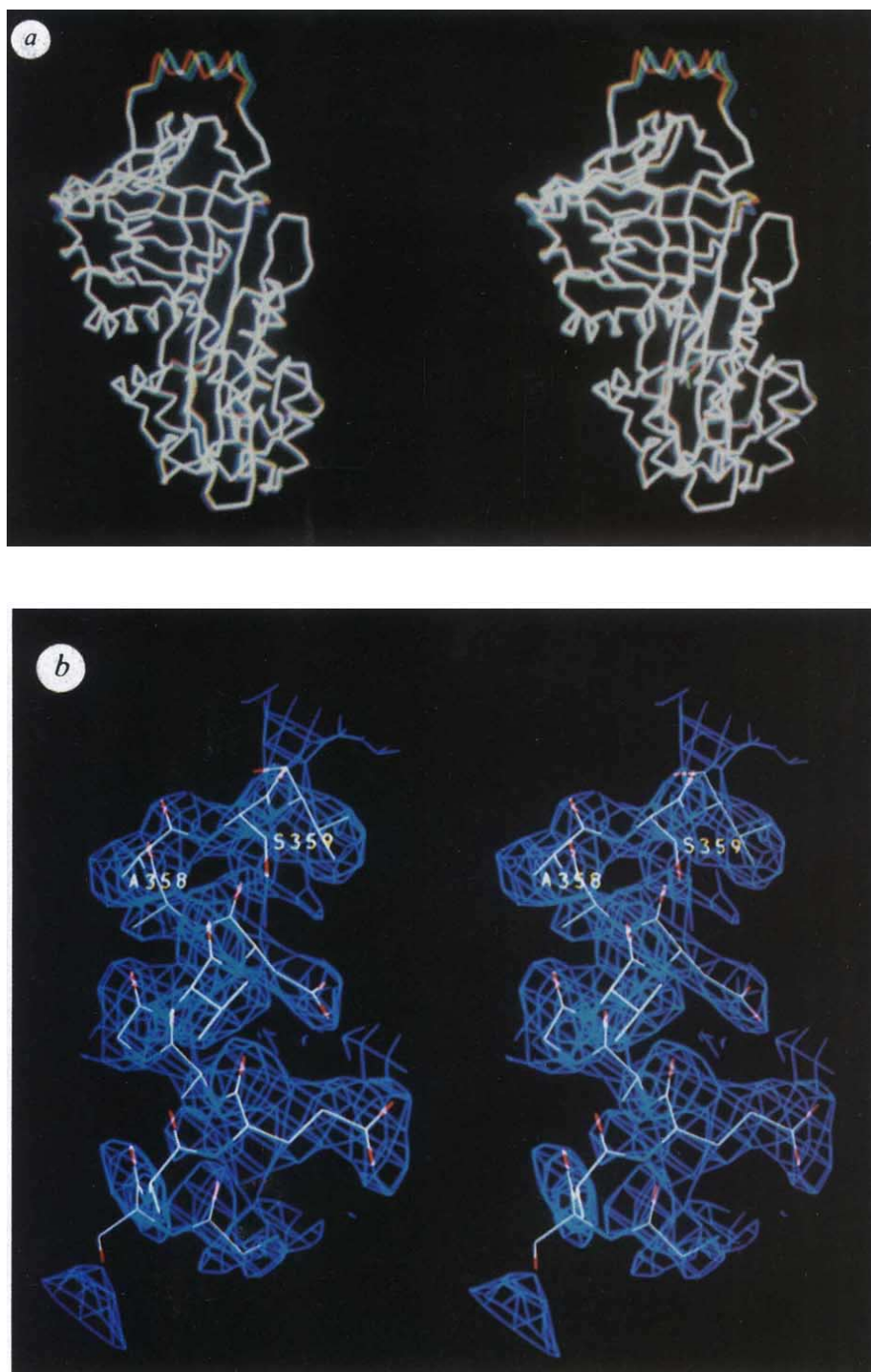


FIG. 3 *a*, A superposition of the α -carbon backbones of three of the four molecules in the unit cell, demonstrating the variability in the position of the helical reactive centre loop (uppermost) with respect to protein core. The electron density in the loop region of the fourth molecule is too weak to allow an unambiguous interpretation, suggesting that the helix is less well ordered than in the other molecules. *b*, An electron density map for the reactive centre helix (residues 349-359) of one of the four molecules in the unit cell calculated with coefficient $(2|F_o| - |F_c|)$, α_c and contoured as the root-mean-squared deviation of the electron density in the unit cell.

be mobile for inhibitory function. The replacement of the P_{12} or P_{10} alanines in antithrombin¹¹ and C1-inhibitor¹² by threonines, serines or prolines, results in complete or partial loss of inhibitory activity, but not of the susceptibility to P_1 - P'_1 cleavage. In effect, these mutations convert the serpins into homologues of ovalbumin which functions as a protease substrate, but not as an inhibitor. This suggests that as a result of the greater flexibility of the N-terminal stalk, the reactive centre loop of an inhibitory serpin is more mobile than the helical loop in the crystal structure of ovalbumin (Fig. 3a). One possibility is that unfolding of the helix in an inhibitory serpin is accompanied by a linked movement of the N-terminal segment back into the A sheet. Thus the cleaved form, in which the whole loop is incorporated as the fourth strand of the A sheet², may represent one extreme of this movement; the other extreme being the full helical form observed in ovalbumin.

Structural studies of the complexes formed by other families of inhibitors with serine proteases have consistently shown that eight residues of the reactive centre loop (P_5 - P'_3) take up an extended conformation^{13,14}. If the helical conformation of the reactive centre loop is a general feature of the serpin family, these observations strongly suggest that for the protease to interact with the inhibitor the helix must be fully unfolded. Solution studies have demonstrated that for an isolated helix, the energy change for the helix to random coil transition is only of the order of 2 kcal mol^{-1} (ref. 15). As the loop helix in ovalbumin is to a large extent structurally isolated, the overall energy required for unfolding is likely to be relatively small. Such a conformational change from an α helix to an extended strand is atypical of protein-protein recognition. Comparison of the structures of free and complexed forms of other protease inhibitors show that although some flexibility is required to facilitate tight binding¹³, larger conformational changes are unusual. Nevertheless the P_1 - P'_1 site of the helical ovalbumin structure can easily conform to the active centre of serine proteases as shown by the catalytic cleavage that occurs at this site with pancreatic elastase¹⁰. If this site is readily accessible for cleavage it is likely that it would be available for putative inhibition.

In conclusion, we believe that the helical reactive centre homologue of intact ovalbumin provides a valid—but stationary—model for the mobile reactive centres of the inhibitory serpins, suggesting an intriguing way in which the serpins interact with their target proteases. □

Received 22 March; accepted 18 June 1990.

- Hunt, L. T. & Dayhoff, M. O. *Biochem. biophys. Res. Commun.* **95**, 864-871 (1980).
- Löbermann, H., Tokuko, R., Deisenhofer, J. & Huber, R. *J. molec. Biol.* **177**, 531-556 (1984).
- Huber, R. & Carrell, R. W. *Biochemistry* **28**, 8951-8966 (1989).
- Wright, H. T., Qian, H. X. & Huber, R. *J. molec. Biol.* **213**, 513-518 (1990).
- Carrell, R. W. & Owen, M. C. *Nature* **317**, 730-732 (1985).
- Pemberton, P. A., Stein, P. E., Pepys, M. B., Potter, J. M. & Carrell, R. W. *Nature* **336**, 257-258 (1988).
- Gettins, P. *J. biol. Chem.* **264**, 3781-3785 (1989).
- Stein, P. E., Tewkesbury, D. A. & Carrell, R. W. *Biochem. J.* **262**, 103-107 (1989).
- Linderström-Lang, K. & Ottesen, M. C. *r. Trav. Lab. Carlsberg* **26**, 403-442 (1949).
- Wright, H. T. *J. biol. Chem.* **259**, 14335-14336 (1984).
- Perry, D. J. & Carrell, R. W. *Molec. Biol. Med.* **6**, 239-243 (1989).
- Levy, N. J., Ramesh, N., Cicardi, M., Harrison, R. A. & Davis, A. E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 265-268 (1990).
- McPhalen, C. A. & James, M. N. G. *Biochemistry* **26**, 261-269 (1987).
- Bode, W., Papamokos, E., Musil, D., Seemueller, U. & Fritz, H. *EMBO J.* **5**, 813-818 (1986).
- Pitts, O. B. *Pure Appl. Chem.* **31**, 227-244 (1972).
- Warner, R. C. in *The Proteins* Vol. 2a (eds Neurath, H. & Bailey, K.) 435-485 (Academic, New York, 1954).
- Goux, W. J. & Venkatasubramanian, P. N. *Biochemistry* **25**, 84-94 (1986).
- Miller, M., Weinstein, J. N. & Wlodawer, A. *J. biol. Chem.* **258**, 5864-5866 (1983).
- Sussmann, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* **A33**, 800-804 (1977).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458-460 (1987).
- Hendrickson, W. A. & Konnert, J. H. in *Computing in Crystallography* (eds Diamond, R., Ramaseshan, S. & Venkatesan, K.) 13.01-13.23 (Indian Institute of Science, Bangalore, 1980).
- Jones, T. A., *J. appl. Crystallogr.* **11**, 268-272 (1978).

ACKNOWLEDGEMENTS. We are grateful to T. Wright and R. Huber for providing the plakalbumin coordinates before publication. We also thank K. Wilson and colleagues at EMBL, Hamburg for help with data collection, T. Creighton and C. Chothia for useful discussion, and the MRC and Wellcome Trust for support. Atomic coordinates for the ovalbumin structure will be deposited in the Brookhaven Protein Data Bank and coordinates for the current model are available from the authors.

nature is available in microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone (____) _____

University
Microfilms
International

Image analysis for all

P.R. Lennard

The availability of user-friendly scientific image analysis software for the Macintosh II has made the application of digital imaging techniques both practical and cost-effective in many areas of research.

ACROSS a wide variety of disciplines, the manipulation of images is an integral part of some aspect of data acquisition, analysis and reporting. Until recently, the leap from scientific image processing using a camera and calipers to digital imaging required a resident programmer and/or expensive and cumbersome hardware and software packages designed for specific applications. With the advent of user-friendly scientific image analysis software for the personal computer, the range of imaging applications accessible to the average laboratory is limited only by the imagination of the researcher.

Setting up

Apple Macintosh II computers, which are useful low-cost platforms for digital imaging, provide power, an intuitive graphical interface, consistency across applications and standardized display and file formats. In fact, if you own a Mac II, or have access to a system, an investment of under \$10 (US) is all that is required to begin image processing.

To start, a copy of the public domain program 'Image' developed by W. Rasband (NIH, Research Services Branch, NIMH, Bethesda, Maryland) can be obtained at no cost from many Mac bulletin boards. Image is a full-function package that can acquire, enhance, measure distances or average density, animate, edit and print images. The \$10 investment will be needed to pay a typesetter service bureau to scan a hardcopy of your image and convert it to a standard Mac PICT or TIFF format file on a floppy disk. This minimal system will not only serve as an introduction to basic image enhancement algorithms and measurement capabilities, but will be sufficient to apply digital imaging to such diverse applications as analysis of electrophoretic gels, quantitative autoradiography, morphometry and kinematics.

Once the user becomes familiar with how digital imaging can be applied to his or her specific needs, the suitability of commercial software for scientific image analysis can be assessed. Of course, if Image is sufficient a \$1,000 to 3,500 (US) software purchase can be avoided. Keep in mind, however, that the commercially available packages not only provide additional enhancement and measurement tools, but also have specific facilities for image management, automated analysis and detailed reporting of results, which may be necessary to make large-scale im-

plementation of your application feasible.

Four of the major software packages (Image Analyst, Automatrix, Inc., Billerica, Massachusetts; TCL-Image, Perceptics Corp., Knoxville, Tennessee; Ultimage, GTFS, Inc., Santa Rosa, California; and Bigimage, 2Ai, Inc., Pasadena, California) have been described in detail and compared to Image¹. For all but the occasional user, hardware to acquire images in the laboratory will also be required. In most cases, a 600–800-line resolution CCD or nuicon videocamera (about \$750) and an 8-bit video digitizing board for the Mac (about \$1,000) are required. If your application generates hardcopy (for example, TEM or gels), a lower-cost alternative to video acquisition is an 8-bit, 300 d.p.i. scanner (prices start at about \$600).

Tools, techniques and tips

The first steps in an image analysis application are the acquisition and, when necessary, enhancement of the image of interest. Newcomers to digital imaging will be confronted with a bewildering array of enhancement algorithms that range from simple optimization of contrast, smoothing, sharpening and edge detection, to much less intuitive tools, such as arithmetic/logical operators and Fourier domain operations. It may be necessary to spend time experimenting in order to establish the sequence of enhancement operations that produces the best results with your images and acquisition system. A useful technique when working with videomicroscopy and fluorescently-labelled probes is real-time frame averaging to reduce noise (see Fig. 1) or frame summation to improve contrast and image quality in low-light situations.

Image acquisition and enhancement is followed by segmentation, in which features of interest are extracted from background. Segmentation is a necessary prelude to quantification. The process may be as simple as manually positioning a selection box around an area from which you wish to make a density measurement. Image provides eight different curve fitting methods for generating density calibration curves against radiation or optical density standards. At the other end of the spectrum are complex segmentation tools such as the Image Analyst normalized correlation function, which can be used to find the location of a spe-

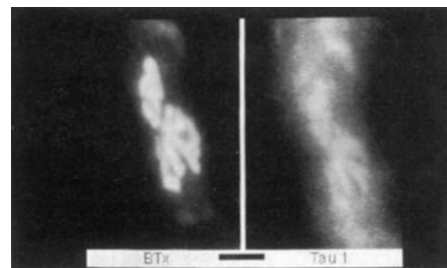


FIG. 1 Macintosh screen displaying the motor endplate from adult rat soleus muscle. Each panel was acquired with 'Image' and results from the averaging of 16 video frames. Left panel: distribution of acetylcholine receptors stained with rhodamine-conjugated alpha bungarotoxin; right panel: distribution of microtubule-associated protein, Tau, in the preterminal axon of the same endplate (stained with a monoclonal antibody and visualized with a secondary antibody conjugated to fluorescein isothiocyanate). (Courtesy of A.W. English.)

cific element of any shape by stepping a search window across the image and finding the best fit to a user-defined template.

Mapping muscle

One of the most important segmentation procedures for scientific image processing is connectivity analysis, which divides images into groups of touching pixels called 'blobs' and constructs a hierarchical list of each blob's perimeter and position within the image. This technique has been used to map muscle architecture and to automate histochemical fibre typing².

After video images of the sections have been digitized, analysis is a two-step procedure. First, the overall shape and area of the muscle, as well as the number, size and shape of each fibre within the muscle, are determined with the aid of Image Analyst's thresholding and connectivity features. The blob-tool is used to recognize each fibre as a separate element, and to mark automatically its centroid with an 'X' (see Fig. 2). This marking provides a convenient interactive way to fine tune thresholding so that each fibre is properly recognized.

Once the blob-tool recognizes each fibre, it is simply a matter of a mouse click to obtain the total number of fibres and the size, shape and boundary points of each individual fibre. These data, along with means and standard deviations for each of the 52 blob parameters calculated by Image Analyst, are saved as a text file, which can be imported to a statistical package for further analysis.

After the size, shape and location of each fibre has been determined, it is necessary to use a second step to sort each fibre into one of four histochemical muscle fibre classes. This is often accomplished by manually locating an individual fibre in three serial sections, each stained for a different histochemical activity. The fibre is then subjectively ranked as showing high, intermediate or low staining intensity in each of the three sections. This three-stain profile, along with the fibre diameter, is used to assign the fibre to one of the four classes.

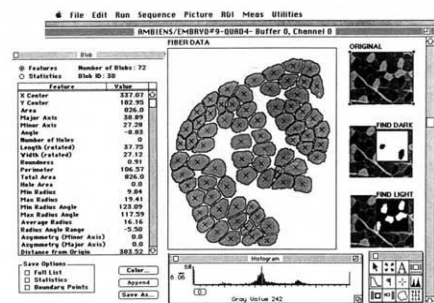


FIG. 2 'Screen dump' printed on an Apple laserwriter of a typical user-interface when fibre typing with 'Image Analyst'. The cursor has been used to select one of the blobs in the FIBRE DATA window (profile in centre of window without black outline) causing its measured parameters to appear to the left of the screen. In the FIND LIGHT window, thresholding was followed by an erosion/dilation, which had the effect of breaking small bridges between adjacent large profiles and leaving open spaces.

Using this technique to rank the several thousand fibres in a muscle cross-section becomes a monumental undertaking and, consequently, less accurate sampling strategies are usually favoured. With Image Analyst, the entire process can be accomplished in a matter of hours. A muscle section stained for alkaline-stable myofibrillar ATPase is acquired and the digital image is further processed by thresholding or density slicing into three sub-images so that connectivity analysis can be performed separately on light, intermediate and darkly stained fibres. This procedure is repeated for the next two serial sections, one stained with NADH-TR (NADH-tetrazolium reductase), the other with mitochondrial alpha-glycerophosphate dehydrogenase. Thus, a total of nine image and nine data files are generated. Final fibre typing can be accomplished by layering the appropriate subset of images. Alternatively, all the data files can be imported to an analysis package where a simple macro can be used to sort and classify fibres automatically.

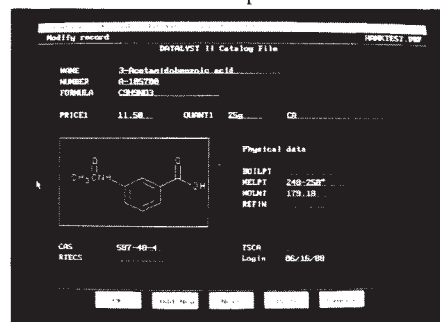
The addition of image processing capabilities to existing computers has provided a low-cost solution to improving data acquisition, analysis and the maintenance of a coherent image database.

Dealing with data

The Scientific Computing and Automation Conference in Philadelphia, Pennsylvania, 18–20 September, will feature software for numeric computation and curve-fitting applications.

The Carcinogenicity Information Database of Environmental Substances (CIDES), developed by Technical Database Services with support from the National Institute of Environmental Health Sciences, is a comprehensive source of numeric data from literature in the field on about 1,000 chemical substances of health and environmental concerns (*Reader Service No. 101*). The database contains 40,000 records taken from a variety of studies relating to carcinogenicity and mutagenicity. The datafiles of CIDES are divided into six categories of studies: human epidemiology, whole animal carcinogenesis, transformations in mammalian cell lines and in mammalian tissues and cells, bacterial tests and other assays. In addition, data are provided on the chemical properties of substances relating to transfer and persistence in the environment. CIDES is available online through the Numerica cluster of computerized information services offered by TDS and may be of interest to environmental chemists, toxicologists, cancer researchers and health and safety officers.

Datalyst II — a chemical database management program for the IBM PC or compatible from Hawk Scientific Systems — allows the user economically to organize chemical structures along with text and numeric data (*Reader Service No. 102*). Two versions are available: a \$795 (US) single-user version, or a networked version where the cost per user can be less

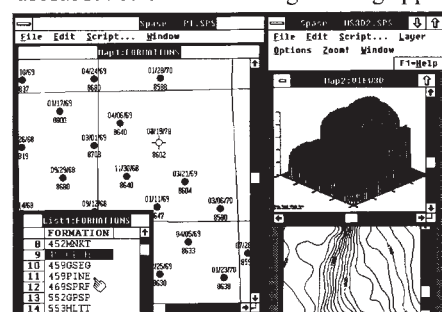


Hawk's PC-based chemical database.

Progress reports, posters and student research papers can now rapidly be assembled in page layout programs, without ever stepping into a darkroom. The down side of adding imaging capabilities to the laboratory Macintosh computer is that you will probably have to wait in line to use it.

than \$500 (US). The program runs on DOS, although a Microsoft Windows version will be available before the end of the year. It uses a graphical user interface with pull-down menus. Screen and report formats are flexible and user-defined, says the company. Data can be searched by substructure as well as by the text and/or numeric data in the database. Complex search queries can be phrased using the simple query language. Simple search queries can, however, be performed with a 'fill-in-the-blanks' query by example. Datalyst II can read and write chemical diagrams in a format compatible with the company's chemical drawing software.

Spase is a graphically-based relational database manager from Geotech Computer Systems that is designed to solve the problems of managing spatial data (*Reader Service No. 103*). The \$2,000 (US) program, which runs on an IBM PC/AT-compatible computer under the Microsoft Windows user interface, is useful for scientific and engineering appli-



Spase — spatial data manager.

cations — geology, geophysics, biology and geography — where some element of spatial data is involved. Data manipulation is performed using an extended version of the industry-standard structured query language which, Geotech says, provides virtually unlimited ways of looking at data. Spase works with all popular contouring, surface modelling and other graphics programs. Coordinate conversion capabilities are provided to allow the user to work with data in latitude-longitude

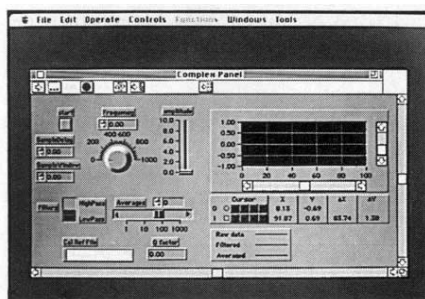
Paul R. Lennard is an Associate Professor in the Department of Biology, Emory University, 1510 Clifton Road, Atlanta, Georgia 30322, USA. For more information, fill in reader service number 100.

- Hollinger, S.J. *ESD: The Electronic System Design Magazine* 26 June (1989).
- Hermanson, J.W., Lennard, P.R. & Takamoto, R.L. *J. Morphol.* **187**, 39–49 (1986).

and cartesian coordinate systems in the same Spase database. Data can be captured directly from hard copy maps and drawings by using digitizers.

An upgraded version 3.3 of New Brunswick Scientific's advanced **fermentation software** provides added flexibility in the monitoring, control and recording of fermentation and cell culture processes (*Reader Service No. 104*). The software, which works with any NBS PC-ready fermentor or cell culture bioreactor, can be used to track all process parameters and control up to four independent bio-processes simultaneously. By setting high and low limits for process parameters, users can be alerted by either an audible or visual alarm to any deviation from these limits. As version 3.3 also produces hard copies of trend graphs, users can incorporate process data into reports.

National Instruments announces the availability of a run-time version of LabVIEW 2, its **Macintosh-based graphical programming software** for instrument control, data acquisition and instrument control solutions (*Reader Service No.*



Run-time only version of LabVIEW 2.

105). Users of the run-time version can load and execute but cannot edit VIs (virtual instruments) — software modules packaged graphically to have the look and feel of an instrument. The new system, priced at \$495 (US), provides, says NI, a cost-effective way to create multiple dedicated workstations running operate-only programs. See the company in booth 609.

DNA/protein programs

If your homology searches fail because genes have diverged, obscuring significant relationships, PATSCAN — the new **software tool for characterizing proteins** from DNASTAR — may provide the answer (*Reader Service No. 106*). PATSCAN can be used to compare a protein sequence to a library of 2,026 diagnostic patterns derived from 10,665 entries of the SWISS-PROT protein sequence database. These patterns encapsulate family characteristics of related proteins in the database and, says DNASTAR, minimize the effect of individual sequence variation. Running on IBM/COMPAQ-compatible micro-computers, including PS/2s, PAT-SCAN

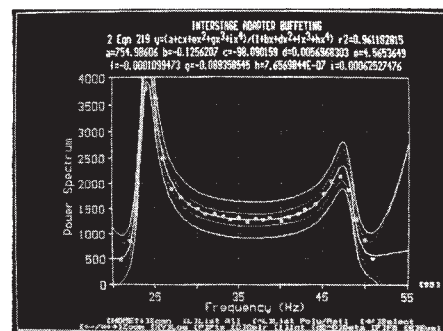
sells for \$500 (US) or \$250 (US) if users have purchased other software from the company.

Busch and Lucas is shipping four new programs running on IBM PC/XT/AT and PS/2 computers or compatible systems under MS-DOS, which are designed to ease you through the **DNA data handling process** (*Reader Service No. 107*). The OLIGO oligonucleotide database program is designed to scan DNA sequences for the occurrence of oligonucleotide sequences. This handy software tool allows users to check oligonucleotide specificities before synthesis. SeqEdit is a nucleotide sequence editor that, in addition to the usual editor features like editing and block operations, can search for subsequences, convert between EMBL, PC-Gene, FlexP and ASCII file formats, translate nucleotide sequences into amino acid sequences, and perform DNA sequence complementing and inverting. For estimating the length of DNA fragments in agarose gels, the company has developed Basepair. Finally, DotMatrix is designed for homology searches in two DNA sequences by graphical means and features a special 'noise' reduction algorithm to make plots more informative.

Data handling

P-Fit is the new **non-linear curve-fitting program** from Biosoft that features data handling, data transformation, equation solving and function evaluation capabilities (*Reader Service No. 108*). Users can choose from the equations already installed, including three four-parameter sigmoid models, linear regression, Michaelis-Menten, 1st-order kinetics, binding and competition, Hill equation, mono- and poly-exponential growth and decay or, alternatively, users can add their own equations using the equation editor. P-Fit provides a complete statistical report on each fit, with parameter values for standard errors, 95 per cent confidence limits, goodness-of-fit index, residuals, covariance, Chi-squared, sign-test and run-test. Data can be exported to word processing packages and desktop publishing programs. The software sells for \$199 (US) or £125 (UK) and is available on either a 3.5- or 5.25-inch disk for IBM PCs, PS/2s and compatibles.

According to Jandel Scientific, its scientific and engineering software package called TableCurve provides a rapid means of **finding the best equation to fit data** (*Reader Service No. 109*). The \$395 (US) TableCurve software program is designed to perform one-pass least squares curve fitting to 221 candidate equations in a single automated processing step. Equations are ranked according to 'goodness of fit' (r^2 coefficients). The user can examine graphically any equation by pressing a



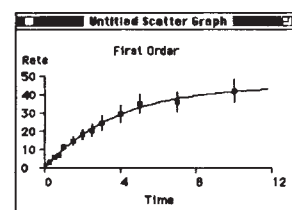
TableCurve — curve-fitting software.

single key and view standard errors, the F-statistic and prediction/confidence tables for the data points. TableCurve runs on IBM PCs or compatibles and can be used to output data to Jandel's SigmaPlot 4.0, Lotus 1-2-3, ASCII and other formats. See booth 307.

DataWave from BrainWave Systems is a new **data acquisition, experimental control and analysis software package** for

ADVERTISEMENTS

New Release



MultiFit 2.0 for Apple Macintosh

Non-Linear Regression
to User-Defined model

Simple, Multiple, Deming and
Robust Linear Regression
Polynomial Regression
and much more...

£100 or US\$190

Orders and enquiries to:

DayComputing

P.O.Box 327, Milton,
Cambridge, CB4 6WL, UK
New Fax +44 223 812726

Scientific Word Processing

MathWord

for users of Microsoft Word

Lab Software

2 Ivy Lane, Broughton, Aylesbury
HP22 5AP England
Tel. 0296 431234 Fax. 0296 397231

ADVERTISEMENTS

ACTIVITY MONITOR

Video tracking system, for water-maze, open field, radial maze etc
In the U.S. 1-800-225-9261

See us at Neuroscience

St. Louis, Oct 29 - Nov 1, Booth #465

HVS IMAGE

Ormond Crescent

Hampton TW12 2TH

England

Tel +44 81 783 1152

Fax +44 81 783 1223



Reader Service No.65

Custom DNA

Purified and Shipped in 48 hours,
Worldwide.

100 ug, \$10.00 a base for the first 15 bases,
\$7.50 for each additional base.
Delivered.

Research Genetics

1-800-533-4363

In the United Kingdom 0-800-89-1393

In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.26

Synthetic Oligonucleotides

Custom made Probes and Primers,
Synthesised, purified and shipped in 3 days

£4.00 per base

(All inclusive)
0.2 µM scale (approx. 500 µg)

Molecular Medicine Unit
King's College School of Medicine & Dentistry
Denmark Hill, London SE5 8RX

Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No.56

DOCUMENT IMAGING

Stem the flood of paper with ATHENE Document Imaging System. Compresses whole filing cabinets onto a few disks; painlessly merges paper and computer information. Uses our own uniquely efficient COLOSSUS text oriented database manager.

Contact:
Paul Gozal
Southdata Ltd
Voysey House
London W4 4PT
081-995 7587/7967/9188

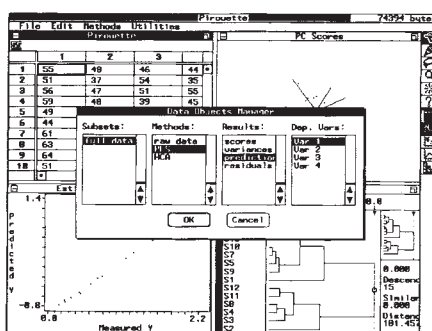
Reader Service No.13

general science applications and research (Reader Service No. 110). The software requires a '286 or '386 AT-based or PS/2 PC, and is, says BrainWave Systems, ideal for applications that require data acquisition and simultaneous data analysis, such as Fast Fourier Transform analysis, waveform extraction and X/Y point plots. Experimental design is performed by the software's experimental editor and experiments can be stored in document-like experimental definition files. External instruments can be controlled using serial communications, digital-to-analog output lines and digital-to-analog converters. Data replay and editing functions allow the user to view information collected during an experiment and edit the data as needed. Data may be replayed as a series of analog waveforms, a stream of data records, or both. Analysed data can be printed, plotted, or saved in standard formats such as ASCII for transfer to other programs. Prices for DataWave start at \$1,995 (US). See booth 1113.

Mathematical means

Blackwell introduces Statistics, the new statistics package for non-statisticians for the Macintosh or PCs running under the Microsoft Windows user interface (Reader Service No. 111). The £129.50 (UK) basic software package can calculate means, variances, standard deviations, coefficient of variation and linear regression. To this, users can add extra modules for t-tests (paired and non-paired), ANOVA (one- and two-way), and non-parametric tests. The basic package plus three extra modules costs £229.50 (UK).

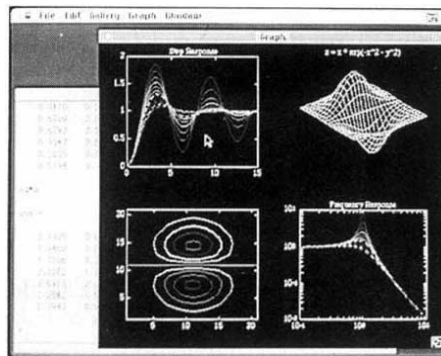
The Pirouette software product from InfoMetrix for IBM PCs and compatible computers provides a comprehensive environment for interpreting multivariate data (Reader Service No. 112). Pattern recognition is performed through the use of principal component analysis and hier-



Pirouette manages multivariate data.

archical clustering. In addition, InfoMetrix says models can be built to either classify samples — using nearest neighbour or SIMCA modelling techniques — or predict a property value. This graphics-orientated package features pull-down menus, dialogue boxes and graphics icons. See booth 518.

MATLAB, the software program for scientific and engineering numeric computation from the MathWorks, is now available for Cray I-S, X-MP and Y-MP supercomputers running under UNICOS and X Windows (Reader Service No. 113). MATLAB combines numerical analysis, matrix computation, signal processing and graphics with an interface that allows users to express problems and solutions in standard maths notation, says the company. It provides users with access to



MATLAB — scientific and engineering software for numeric computation.

linear algebra and matrix algorithms from LINPACK and EISPACK, and provides built-in functions for differential equation solution, polynomial operations, matrix computation, statistics and other numeric techniques. Users are able to edit existing functions or create new ones. Data can be displayed graphically as MATLAB provides 2-D linear, log, semi-log and polar plots, as well as 3-D 'mesh' and contour graphs. Relevant scientific disciplines include applied mathematics, physics, astronomy, chemistry, geophysics, electrical and mechanical engineering. See The Mathworks in booth 400.

The latest Hewlett Packard software module — HP 89531A — for the HP 8452A general purpose UV/visible spectrophotometer adds the capability for multicell enzyme kinetics to the previous single-cell capability (Reader Service No. 114). The software includes calculation parameters for beginning and ending times, and a factor for calculating enzyme activity or substrate quantities. Three calculation methods are provided as standard: zero order, first order and end point. Run parameters can be stored on disk and accessed when necessary; this reduces operator error and set-up times, says HP. The software can evaluate spectral data and data can be transferred to spreadsheet programs such as Lotus 1-2-3 or Excel. See HP in Philadelphia in booth 405.

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.